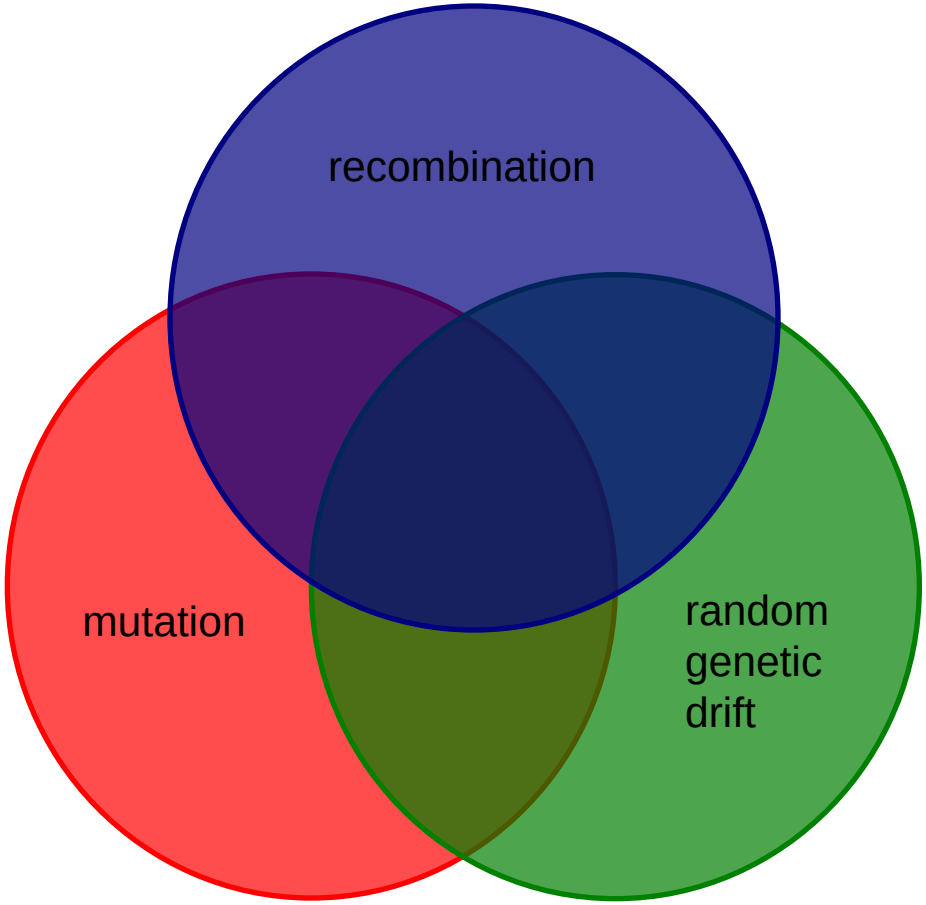
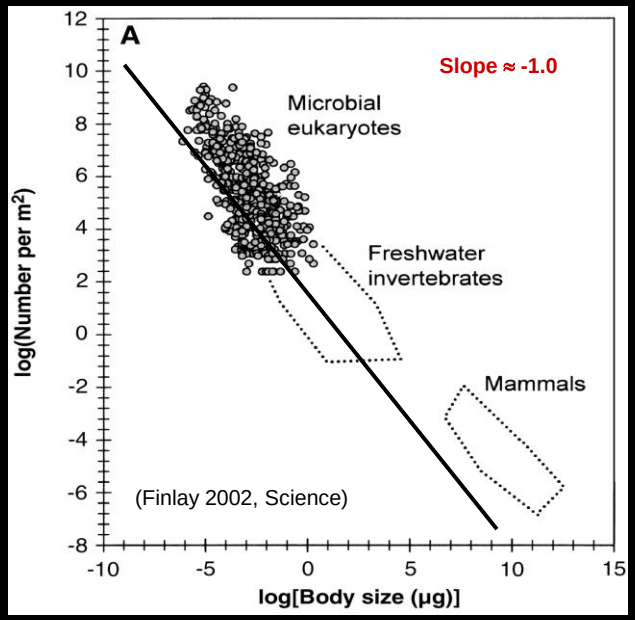
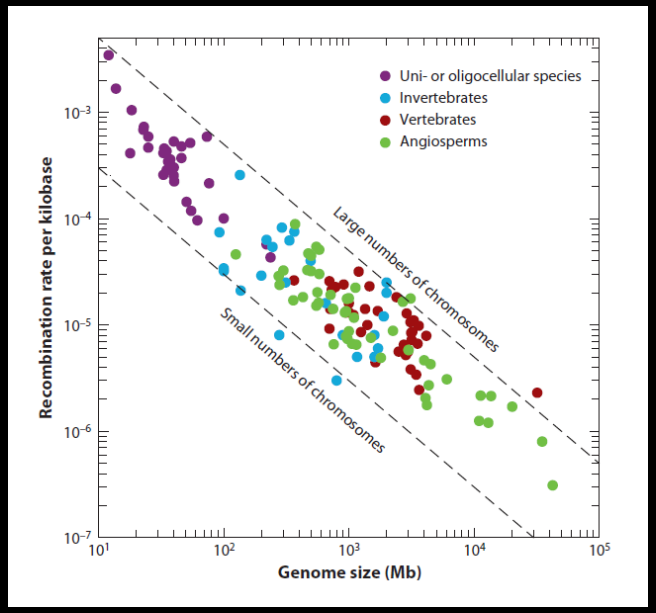
The Population-genetic Environment



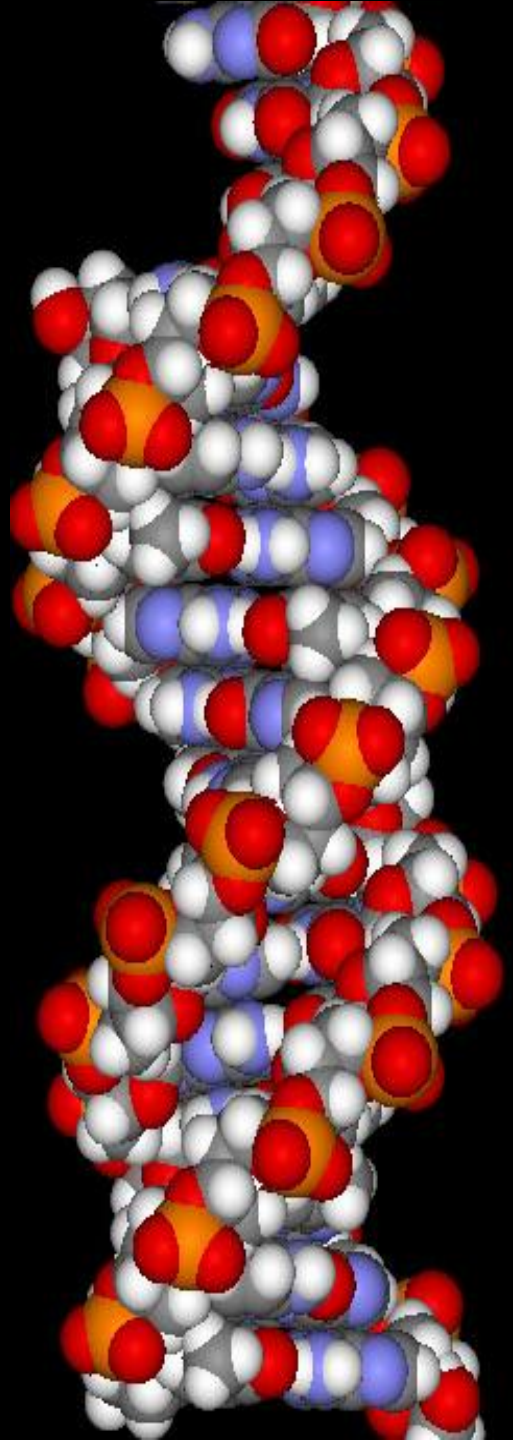
Scaling of Population Size and Recombination Rate with Organism Size



Reduction in absolute population size



Reduced recombination per physical distance



Evolution of the Mutation Rate

- The mutation rate scales across phylogenetic groups, among tissues, and among polymerases within cells.

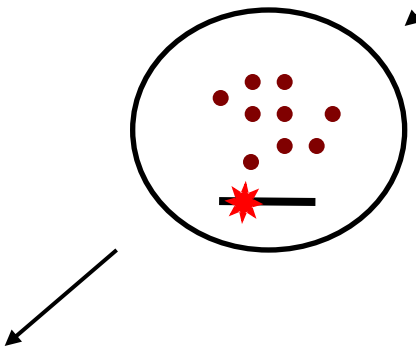
- No evidence that mutation rates have been optimized to maximize the long-term rate of adaptive evolution.

- No evidence that the efficiency of replication has been pushed to the limits of molecular perfection.

- **The Drift Barrier to Mutation-rate Reduction**: Once the selective advantage of lowering the mutation rate is less than the power of drift, $1/(2N_e)$, the mutation rate has reached its minimum possible value.

The Magnitude of Selection Operating to Improve Replication Fidelity

Asexual Populations: the selective disadvantage of a weak mutator allele
= the increase in the genome-wide deleterious mutation rate



ΔU, increase in genome-wide rate of deleterious mutation

S, rate of removal by selection

Excess number of mutations at equilibrium = $\Delta U / s$

↓

X

Effect / mutation = s

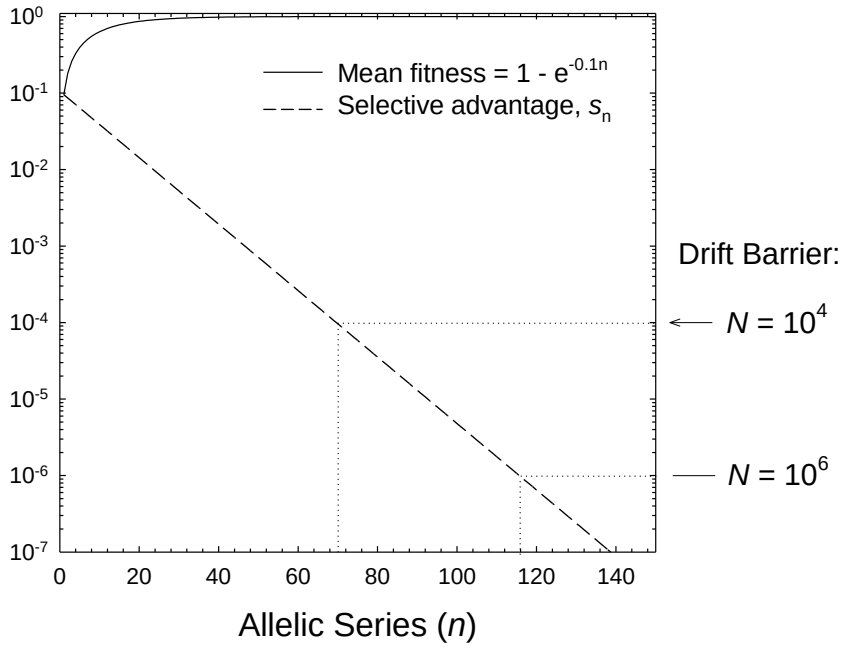
↓

Total effect on fitness = ΔU

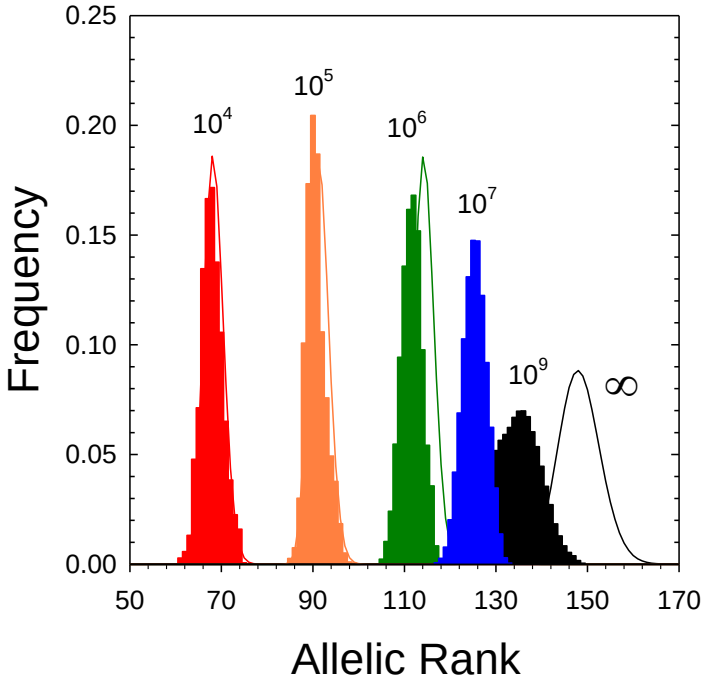
Sexual Populations: the selective disadvantage of a mutator allele is much smaller, $2s \cdot \Delta U$, because recombination prevents the buildup of linked mutations.

The Drift-barrier Hypothesis for a Single Trait

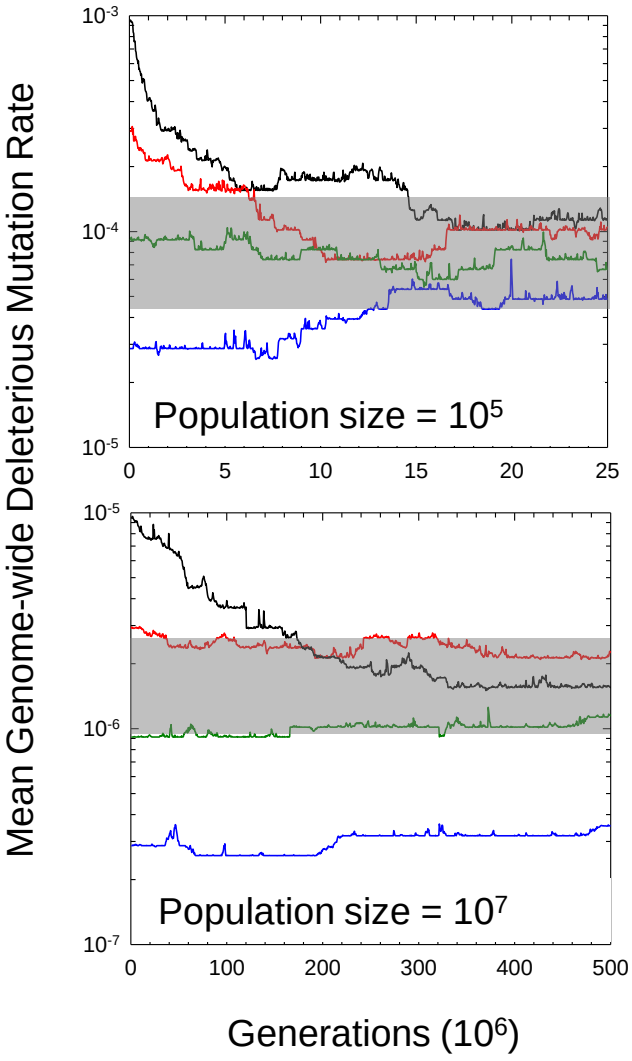
Asymptotically Increasing Perfection in an Allelic Series



Equilibrium Frequencies of Alleles with Increasing Population Sizes



Quasi-equilibrium Mutation Rates Resulting From Deleterious-mutation Load

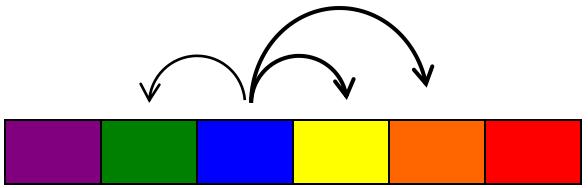


Effective selection for antimutators

DRIFT BARRIER

Biased production of mutators

- Equilibrium mutation rate is inversely proportional to the effective population size.
- Asymmetry of rate of approach to equilibrium.



Mutation-rate classes

The Evolution of Neutrality for the Efficiency of an Enzymatic Function: an inevitable outcome of natural selection (Hartl et al., Genetics, 1985).

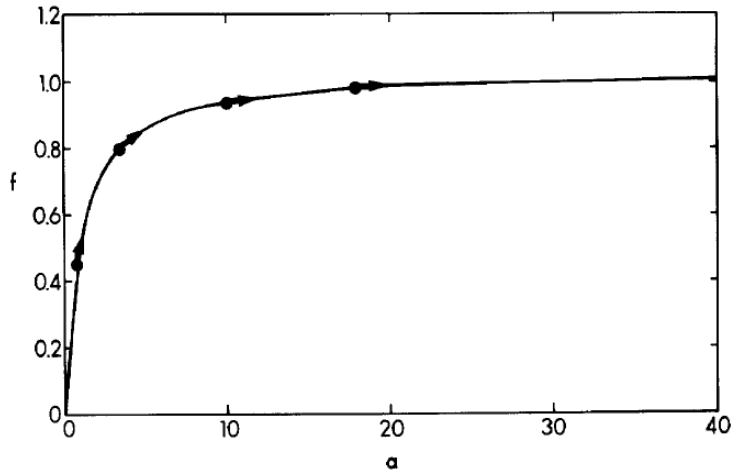


FIGURE 1.—Standardized Michaelis-Menten saturation equation $f(a) = a/(1 + a)$ scaled to equal at $a = 30$.

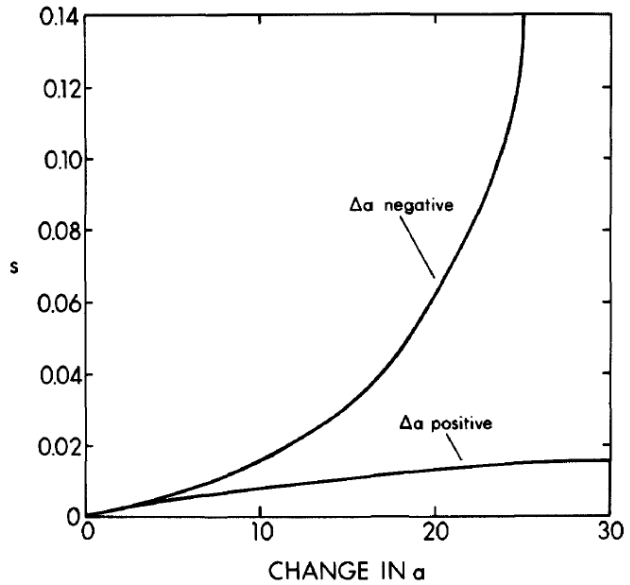
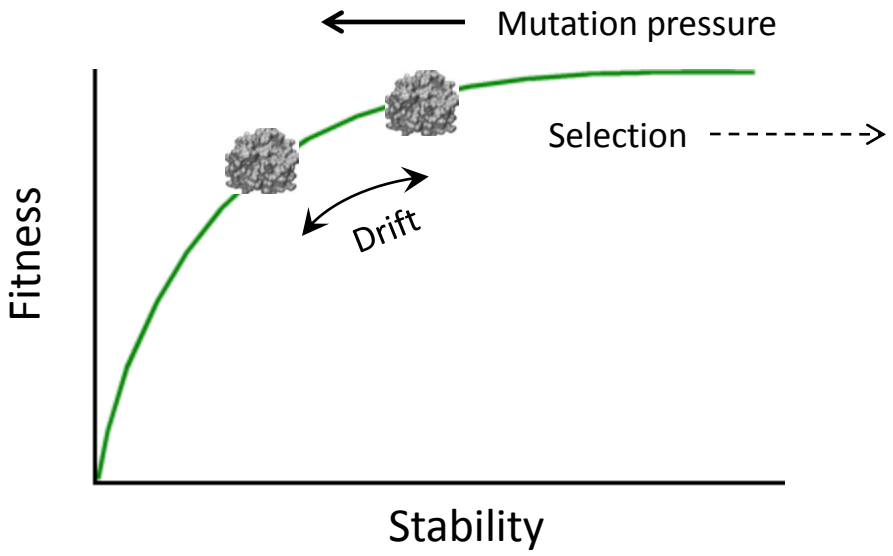


FIGURE 2.—Selection coefficient (s) resulting from a given positive or negative change in enzyme activity (Δa), when initially $a = 30$.

Selection-mutation Balance and the Margin of Protein Stability

- One potential explanation for marginal stability is that overly rigid proteins will compromise protein function.
- However, this argument is inconsistent with observations indicating that proteins engineered to have higher stability often have normal enzyme function.



Drake's (1991) Conjecture:
A Constant Rate of Mutation per Genome per Cell Division in Microbes

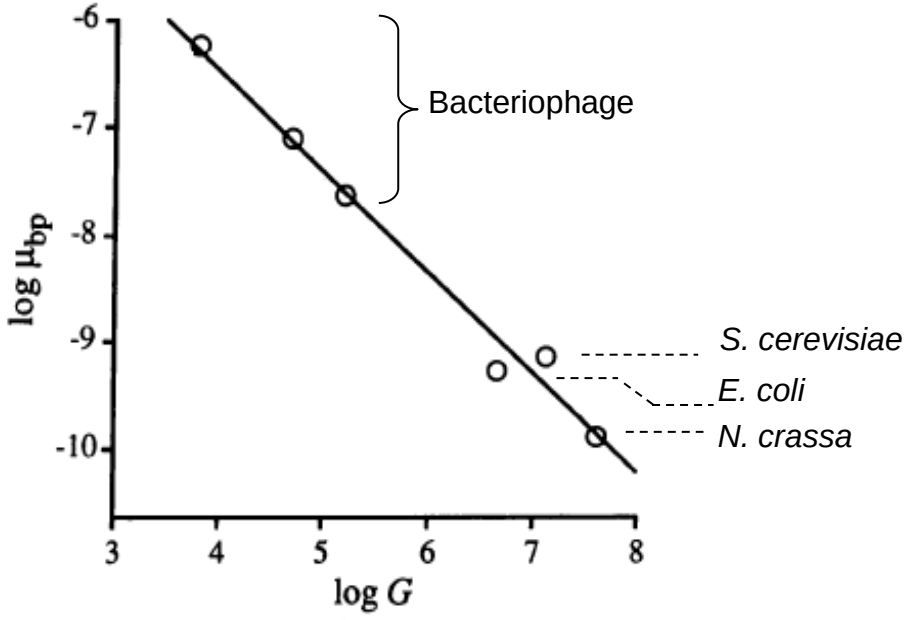
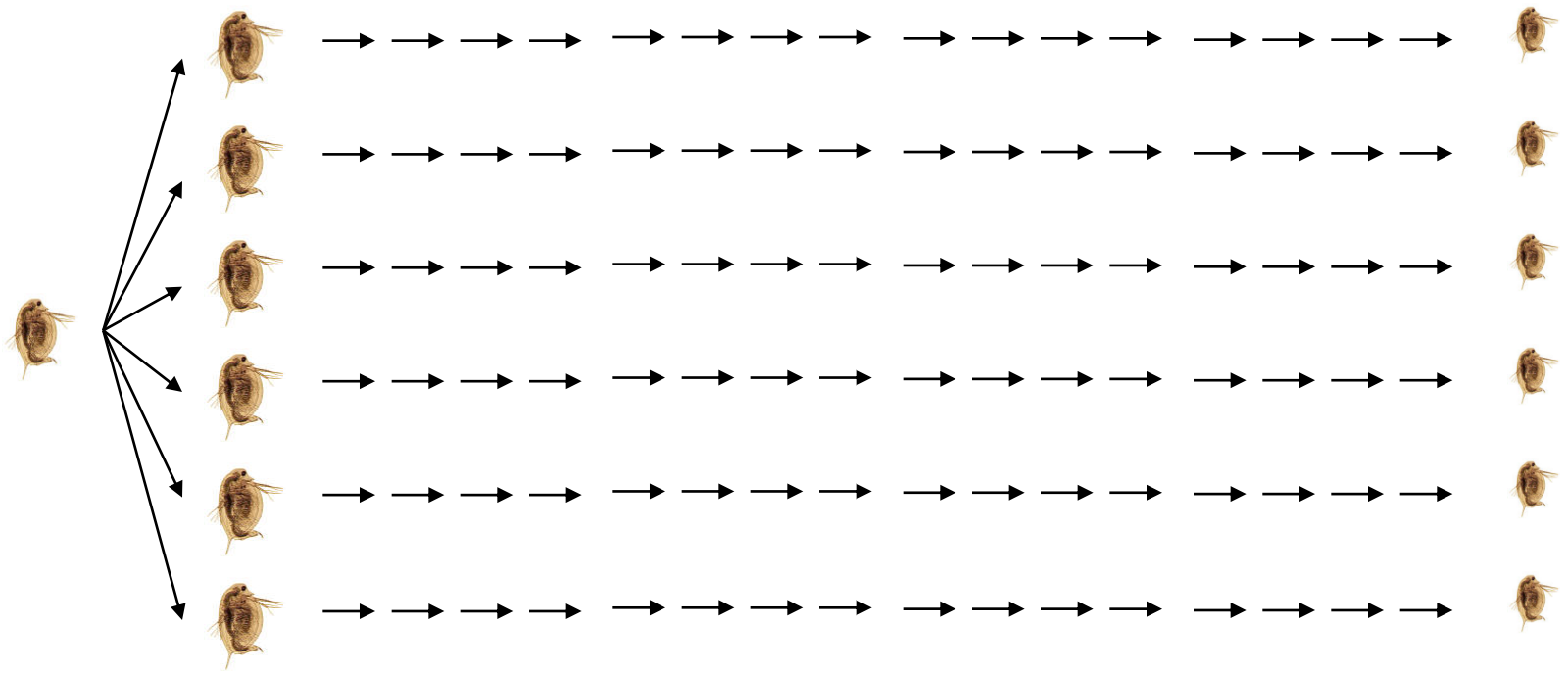


FIG. 1. Average mutation rate μ_{bp} per base pair as a function of genome size G in bp. The logs of the rates for each organism were averaged and all 13 values are included. Phages T2 and T4 were treated as a single organism.

"Because this rate is uniform in such diverse organisms, it is likely to be determined by deep general forces."

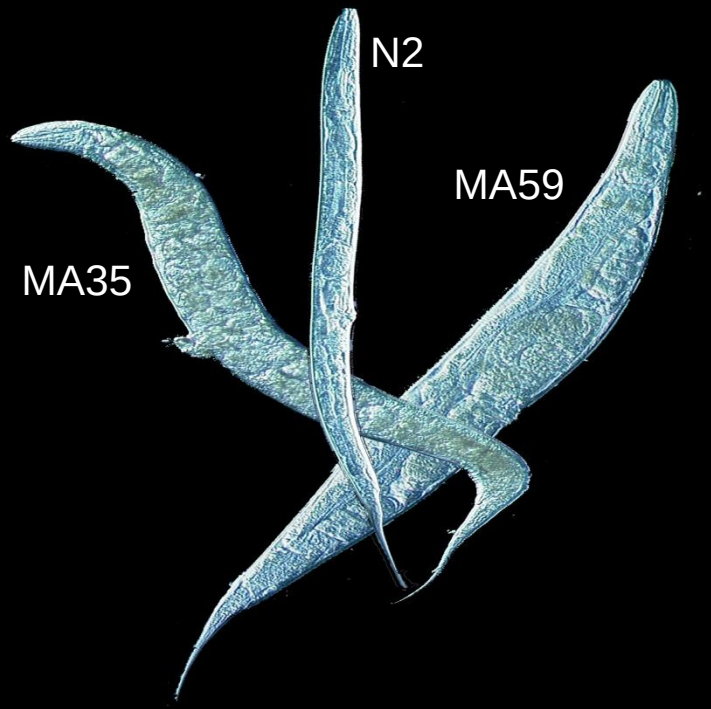
Mutation-accumulation experiment. Starting with a single stem mother, sublines are maintained by single-progeny descent, preventing selection from removing spontaneous mutations. This protocol is continued for hundreds of generations with dozens of lines.



Advantage – essentially no selection bias; allows a genome-wide perspective of the entire molecular mutation profile, from substitutions to large deletion/duplications.

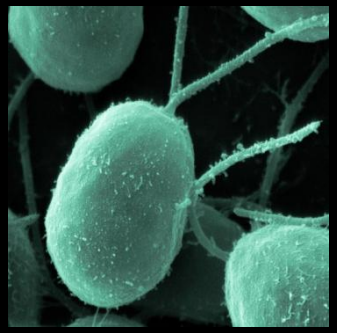
Disadvantage – labor intensive; line / investigator loss.

Extreme Morphological Divergence in MA lines of *C. elegans*



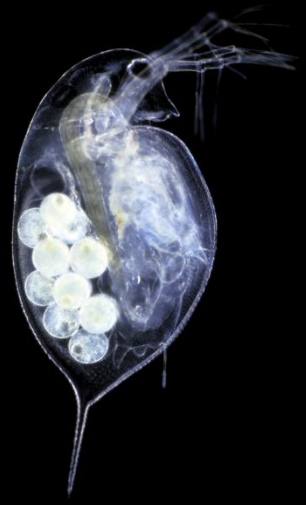
Recent and Current Eukaryotic Targets of Study

Arabidopsis

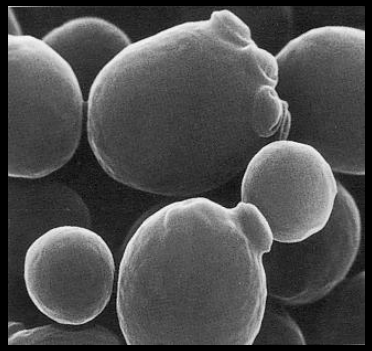


Chlamydomonas

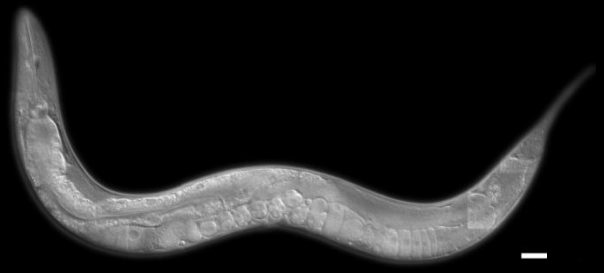
Phaeodactylum



Daphnia



Saccharomyces



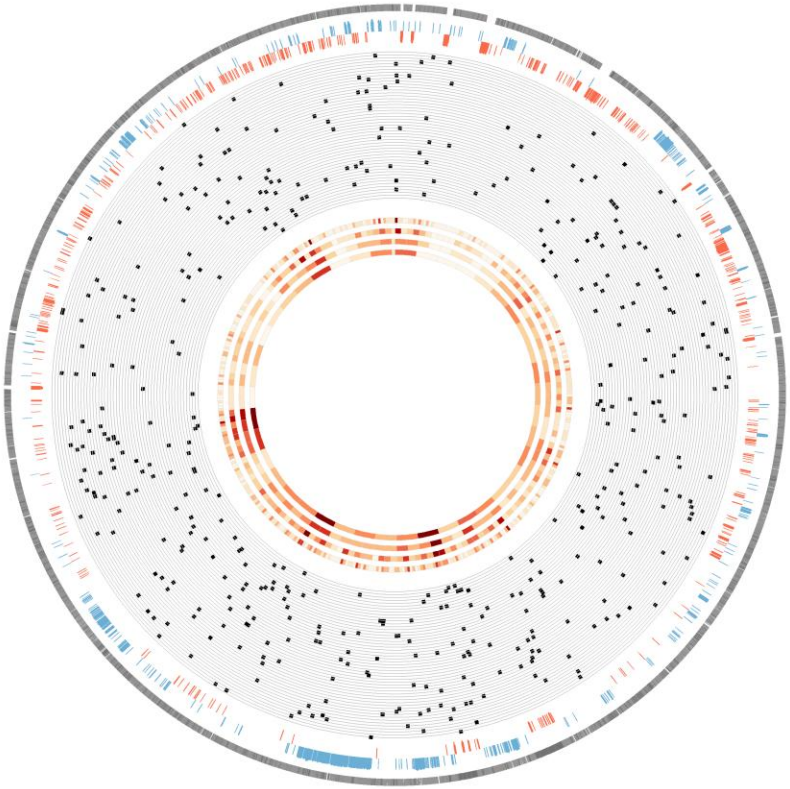
Caenorhabditis

Paramecium

Mutation-accumulation experiments in diverse bacterial species: full range of genome sizes, G:C content, and roles in the environment and pathogenesis.

Phylum	Species	Genome Size (Mb)	Genome G/C (%)	Progress	Significance
Firmicutes	*Bacillus subtilis*	4.2	44	completed	Soil bacterium
Firmicutes	*Escherichia coli*	4.6	51	completed	Food pathogen
Firmicutes	*Mesoplasma florum*	0.8	27	completed	Synthetic cell
Firmicutes	*Staphylococcus epidermidis*	2.6	32	sequencing	Infectious bacteria
Proteobacteria	*Vibrio cholerae*	4.1	48	5600	Cholera
Proteobacteria	*Burkolderia cenocepacia*	7.8	67	4800	Cystic fibrosis
Proteobacteria	*Vibrio fischeri*	4.3	38	4500	Squid symbiont
Proteobacteria	*Pseudomonas fluorescens*	7.1	63	4300	Pathogen control
Proteobacteria	*Caulobacter crescentus*	4.0	67	3000	Synchronized growth
Proteobacteria	*Agrobacterium tumefaciens*	5.7	59	3000	Tumor-inducing bacteria
Proteobacteria	*Rhodobacter sphaeroides*	4.5	68	1900	Phototrophic bacteria
Proteobacteria	*Teredinibacter turnerae*	5.2	51	1700	Mollusk symbiont
Proteobacteria	*Photorhabdus luminescens*	5.7	43	500	Nematode symbiont
Deinococcus	*Deinococcus radiodurans*	3.2	67	4500	Radiation tolerant
Actinobacteria	*Kineococcus radiotolerans*	5.0	74	3500	Radiation tolerant
Euryarchaeota	*Haloferax volcanii*	4.0	66	500	High salt growth
Acidobacteria	*Acidobacterium capsulatum*	4.1	61	500	Growth at low pH
Planctomycete	*Gemmata obscuriglobus*	9.0	67	200	Ammonium oxidation
Firmicutes	*Streptococcus pneumoniae*	2.1	40		Pneumonia
Proteobacteria	*Myxococcus xanthus*	9.1	69		High gene duplications
Proteobacteria	*Serratia proteamaculans*	5.5	55		Pneumonia association
Proteobacteria	*Agrobacterium vitis*	6.3	58		Crown gall disease
Proteobacteria	*Rhizobium sp. NGR234*	6.9	62		Nitrogen fixation
Proteobacteria	*Campylobacter jejuni*	1.8	30		Food contamination
Spirochaetes	*Brachyspira hyodysenteriae*	3.0	27		Swine dysentary
Euryarchaeota	*Methanococcus voltae*	1.9	29		Methanogen
Euryarchaeota	*Methanocaldococcus jannaschii*	1.8	31		Methanogen
Actinobacteria	*Mycobacterium tuberculosis*	4.4	29		Tuberculosis
Cyanobacteria	*Synechococcus elongatus*	2.7	56		Marine carbon fixation
Cyanobacteria	*Synechocystis sp. PCC6803*	4.0	47		Marine carbon fixation

Mutation in Small vs. Large Genomes

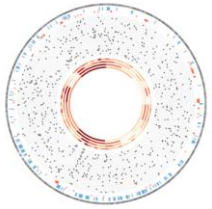


Index:

Outer Rings
- Gene Density
- High G/C Region
- High A/T Region

Intermediate Rings
- Mutations

Inner Rings
- Mutation Density
  - Window Size (1k, 5k, 25k, 100k)



Bacillus subtilis 3610

Genome size: 4,214,598 bp

GC content: 43.5%

50 lines - 450 mutations - 5000 generations

Mutation Rate : 3.27×10^{-10}/site/gen.

Mesoplasma florum L1

Genome size: 793,224 bp

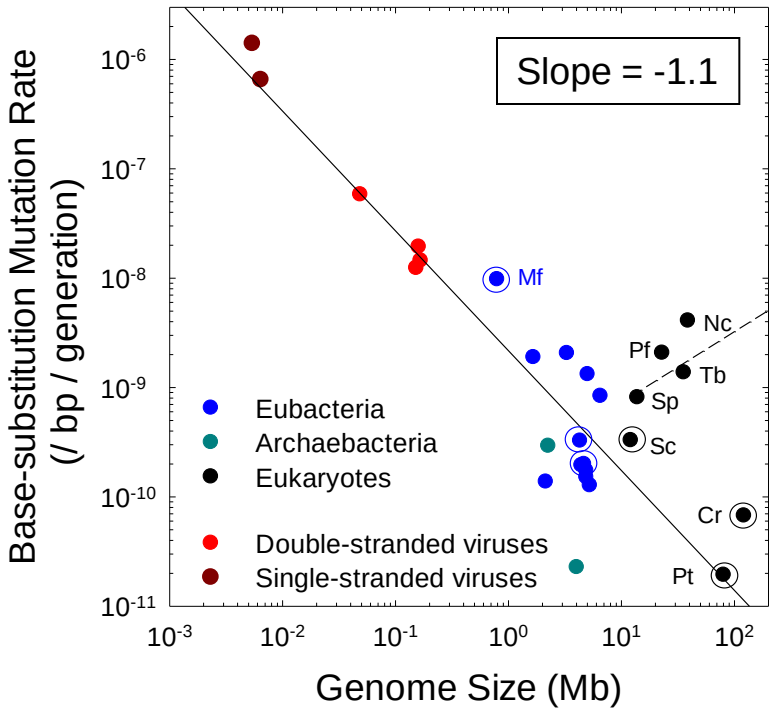
GC content: 27.0%

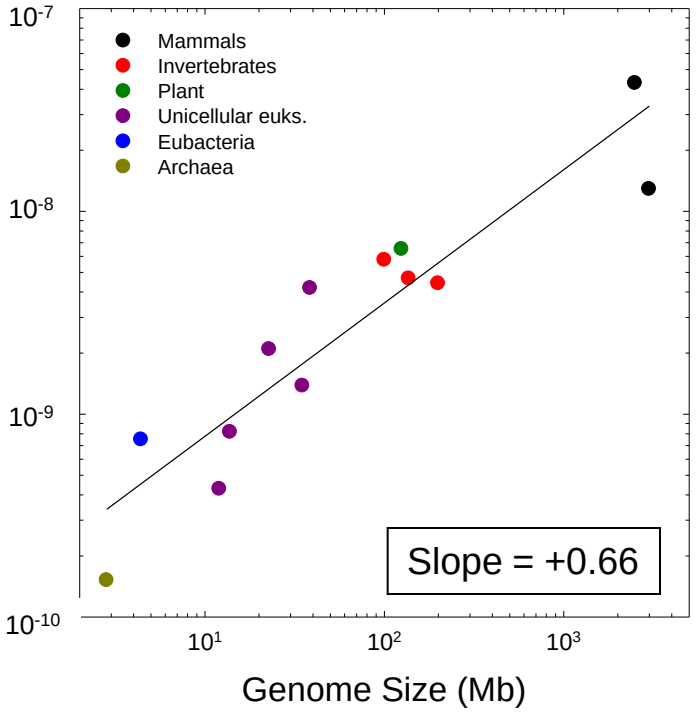
50 lines – 599 mutations - 2000 generations

Mutation Rate : 1.14×10^{-8}/site/gen.

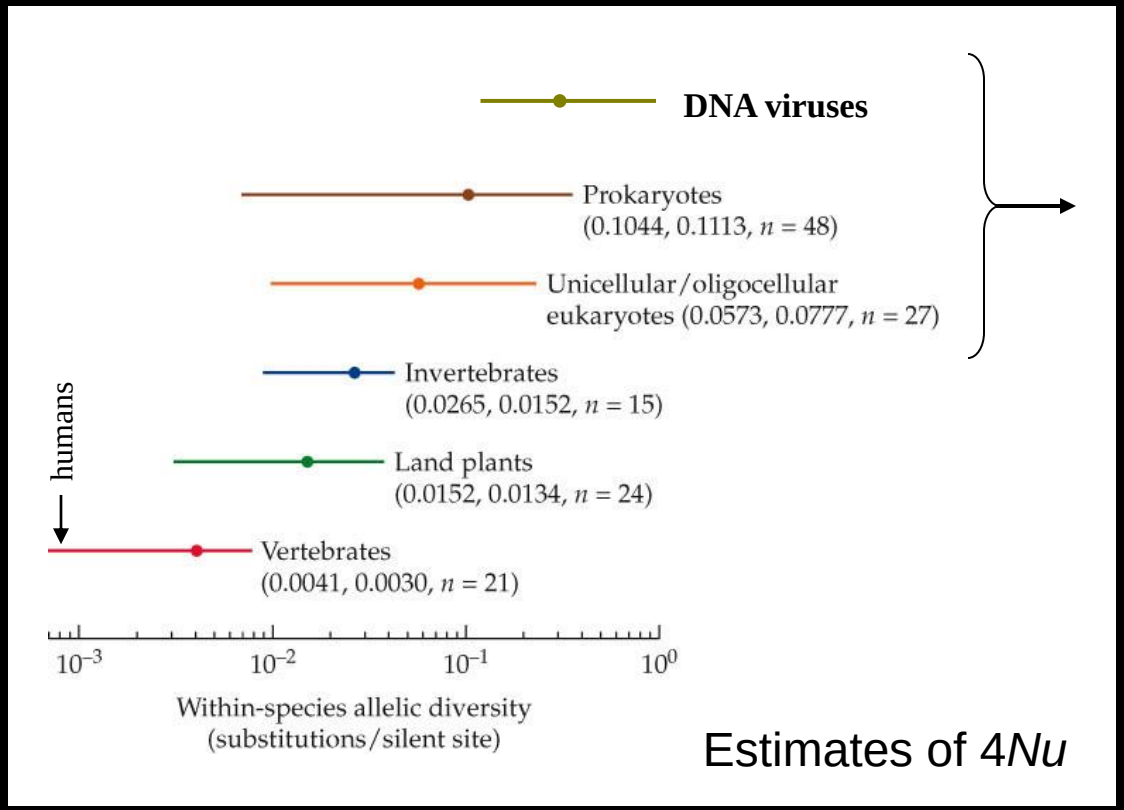
- The *average* number of mutations per genome per generation is roughly constant in **noneukaryotic microbes**, in accordance with Drake's hypothesis.

- The mutation rate per nucleotide site increases with genome size in **eukaryotes**, yielding a dramatic increase in the genome-wide mutation rate per generation.

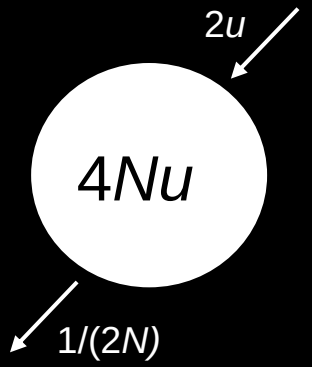




Estimates of the ratio of the power of mutation ($2u$) to the power of random genetic drift ($1/2N$) from standing population-level nucleotide heterozygosity at silent sites.



Estimates of $4Nu$

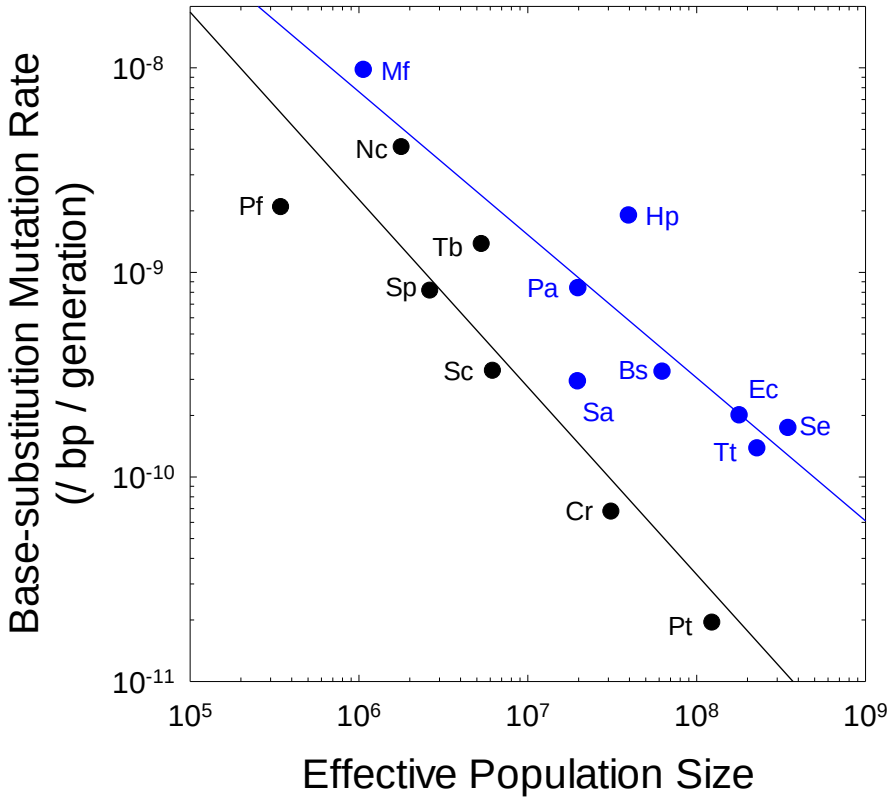


At equilibrium, average allelic divergence at neutral sites =

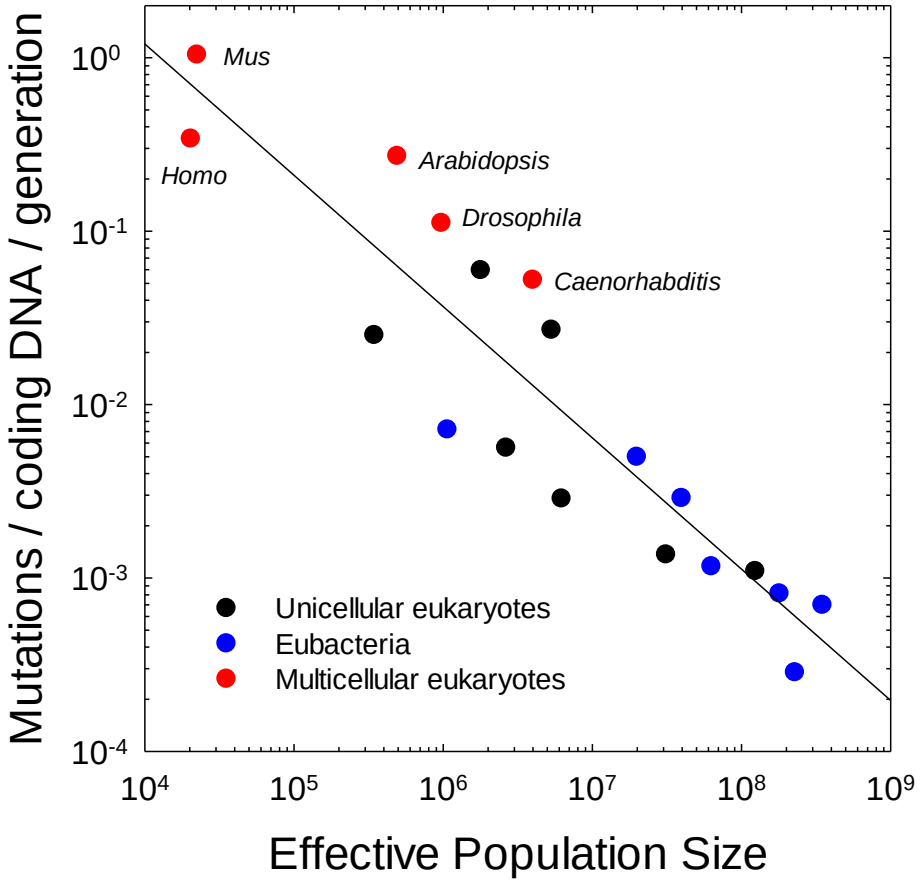
ratio of the power of mutation to the power of random genetic drift.

The Mutation Rate / Nucleotide Site Is Inversely Proportional to the Average Effective Population Size of a Lineage

For a given magnitude of genetic drift, selection is capable of driving the mutation rate down further in eukaryotes.

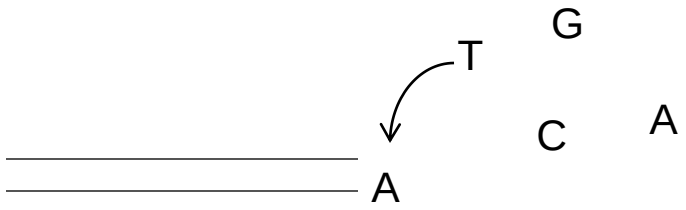


A Universal Inverse Scaling Between the Genome-wide Deleterious Mutation Rate and N_e Across the Tree of Life

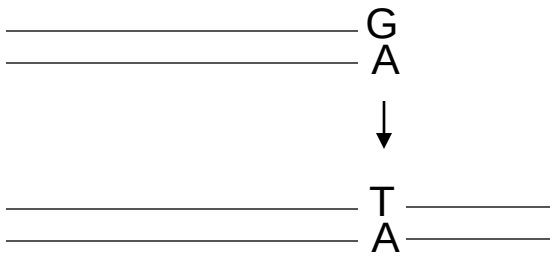


The Three Molecular Lines of Defense Against Mutation

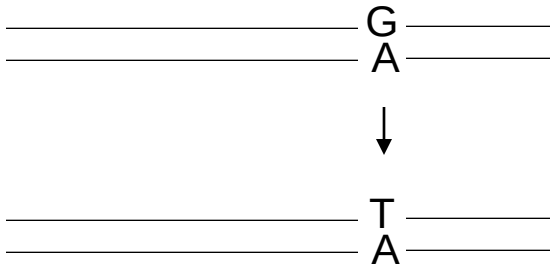
1) Polymerase base-incorporation fidelity:



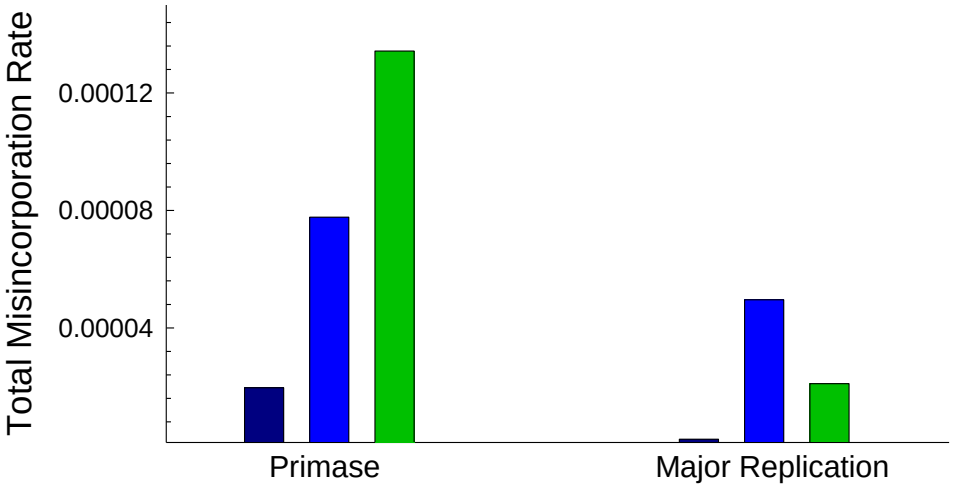
2) Polymerase proofreading:

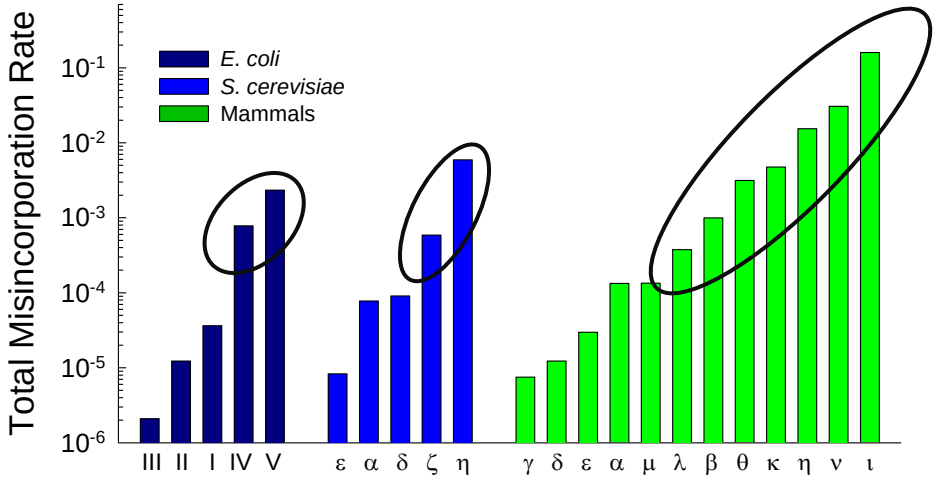


3) Post-replicative mismatch repair:



Polymerase Error Rates Are Magnified in Eukaryotes and in Enzymes Involved in Fewer Nucleotide Transactions

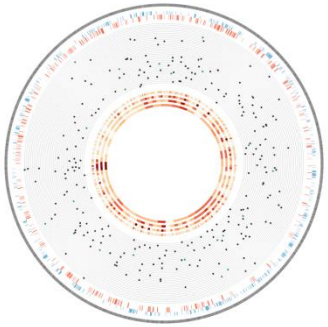




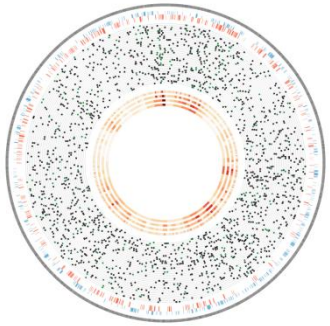
Polymerases used in DNA repair are highly error prone, consistent with the drift hypothesis: enzymes involved in fewer nucleotide transactions experience less selection for fidelity.

Measuring the Efficiency of the Mismatch-repair Pathway

- 150x increase in the mutation rate in *E. coli*: mismatch repair normally corrects ~99.3% of replication errors.

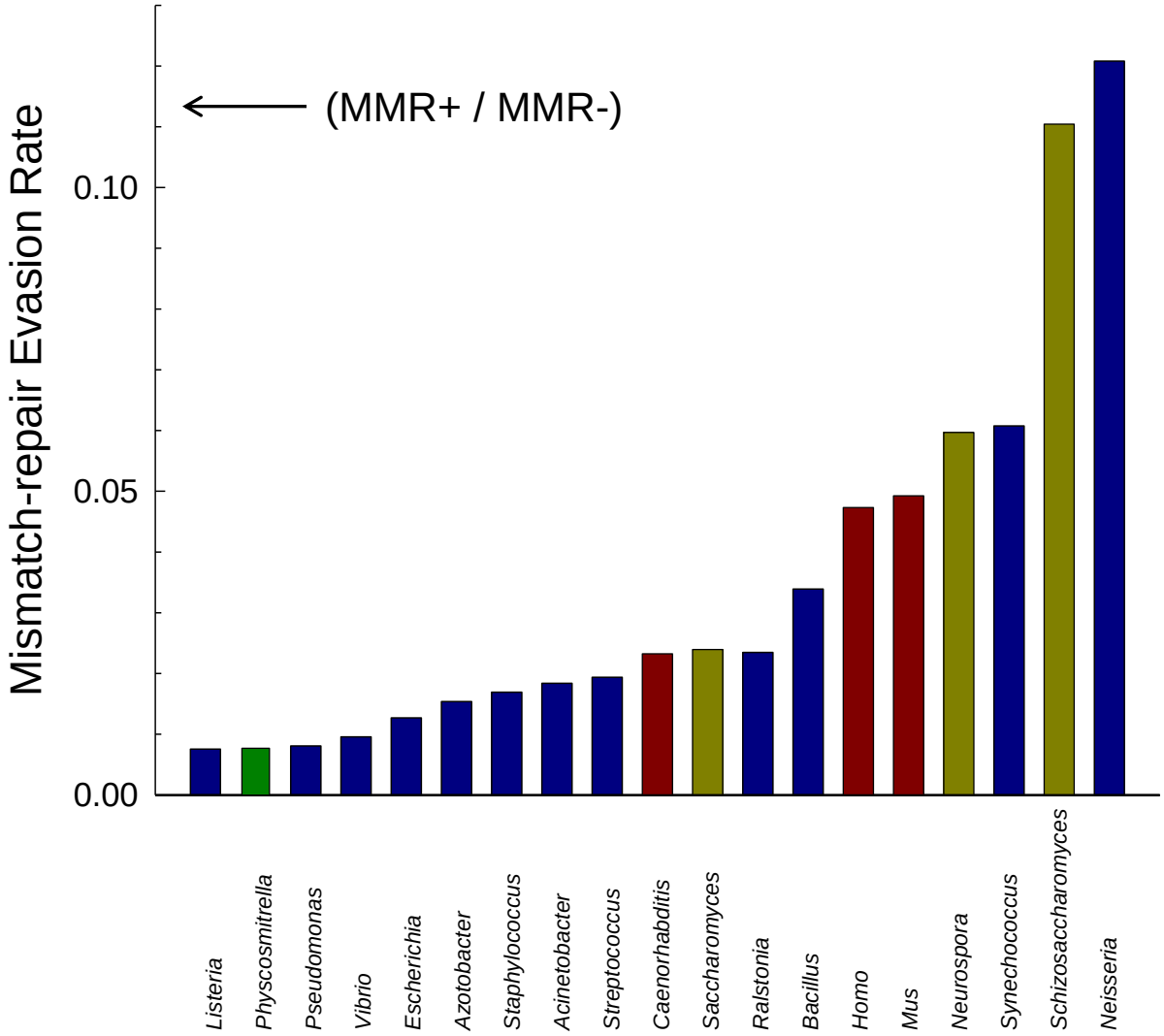


Wild-type
59 lines - 254 mutations - 6000 generations
Mutation rate: 2.16×10^{-10}/site/gen.



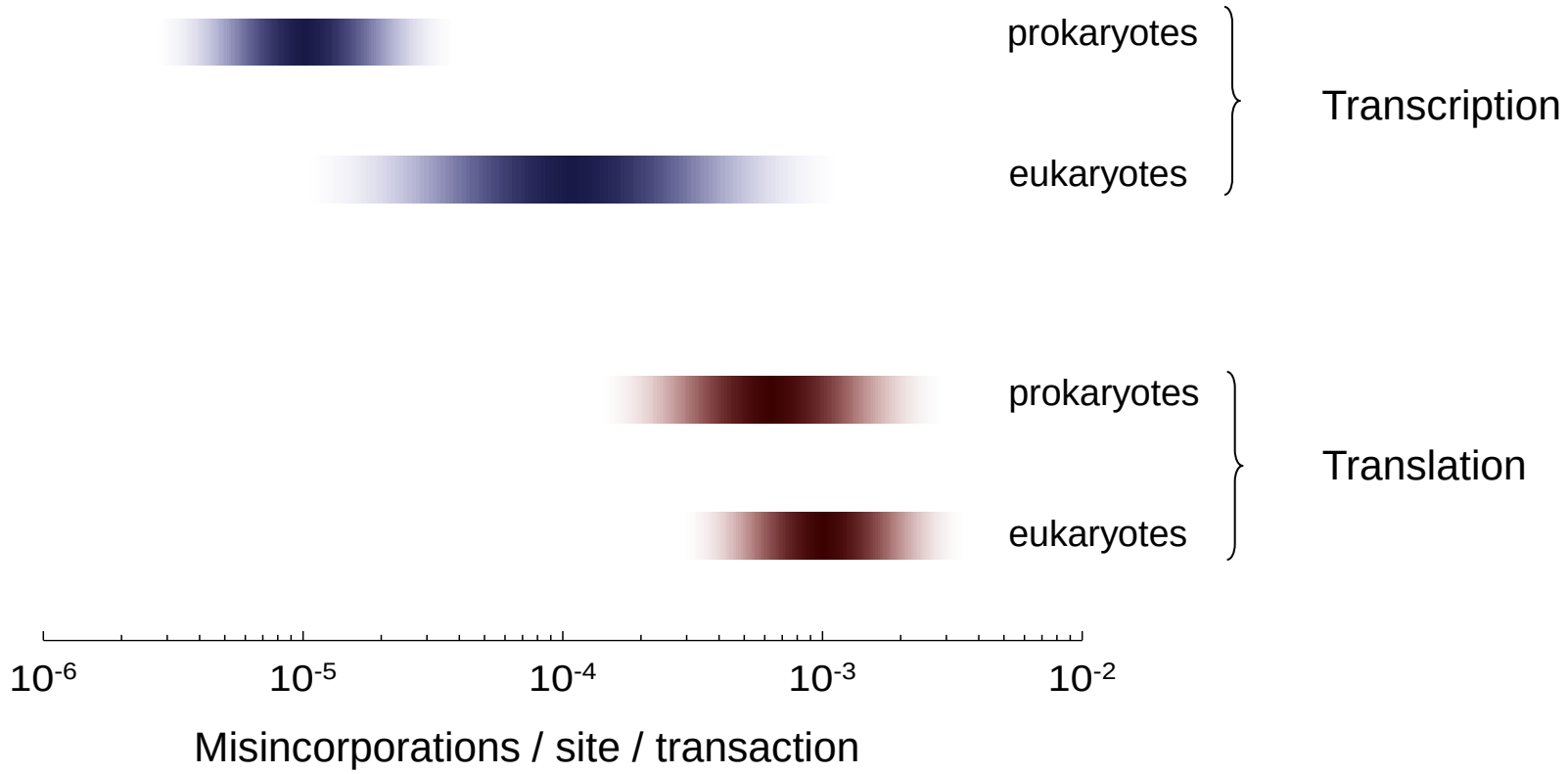
Mismatch-repair Defective
34 lines - 1931 mutations - 375 generations
Mutation rate: 3.26×10^{-8}/site/gen.

Mismatch Repair: the third line of defense is much less efficient than the polymerization and proof-reading steps



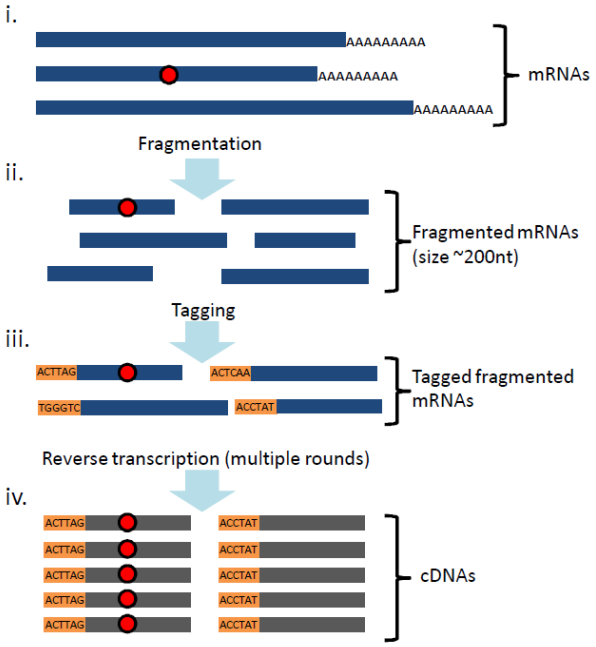
- The fidelity of this downstream repair pathway is >100x lower than that for the upstream polymerase, consistent with the drift hypothesis.

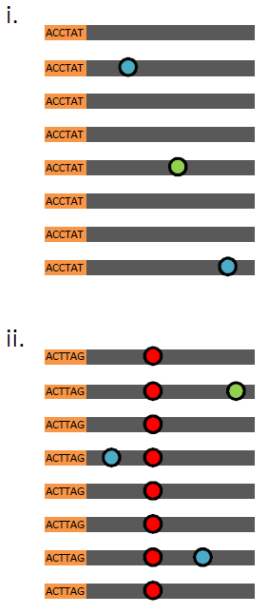
Transcription and Translation Error Rates Are Thought to be Orders of Magnitude Higher Than Replication Error Rates

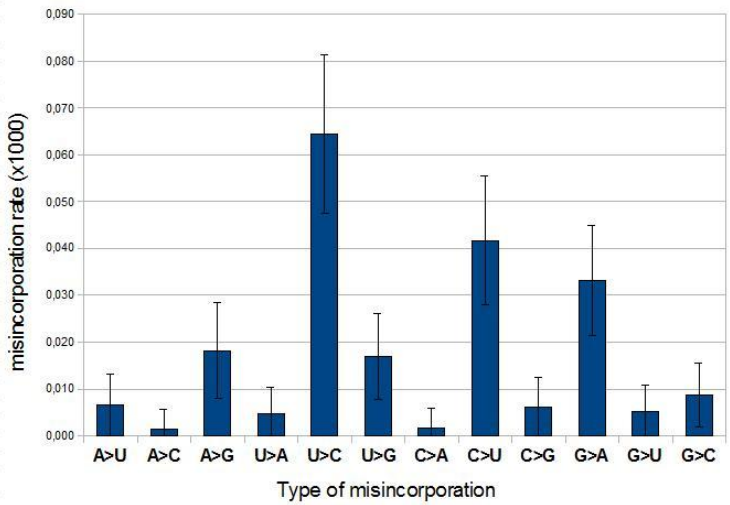


Because products of transcription and translation are more transient than inherited germ-line mutations, selection to reduce error rates at these levels is less efficient.

Direct Estimation of Genome-wide Transcription-error Rates From mRNAs

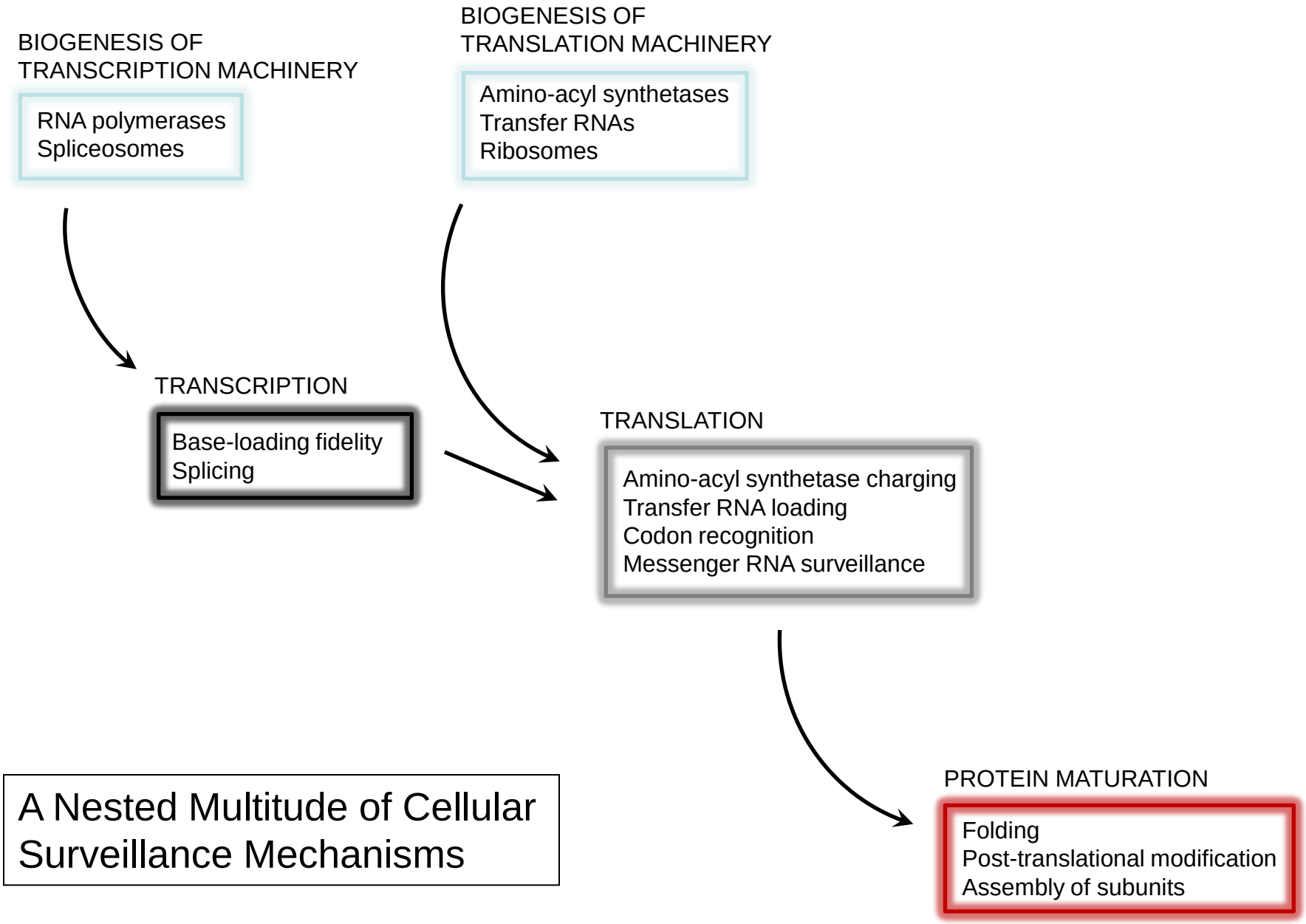






Work with J.-F. Gout and W. K. Thomas

- The transcription-error rate in *C. elegans* is ~10^{-5} per site, which is ~2500x the genomic mutation rate.
- ~3% of transcripts contain errors.
- As the translation error is likely even higher, probably ~10% of proteins contain errors.

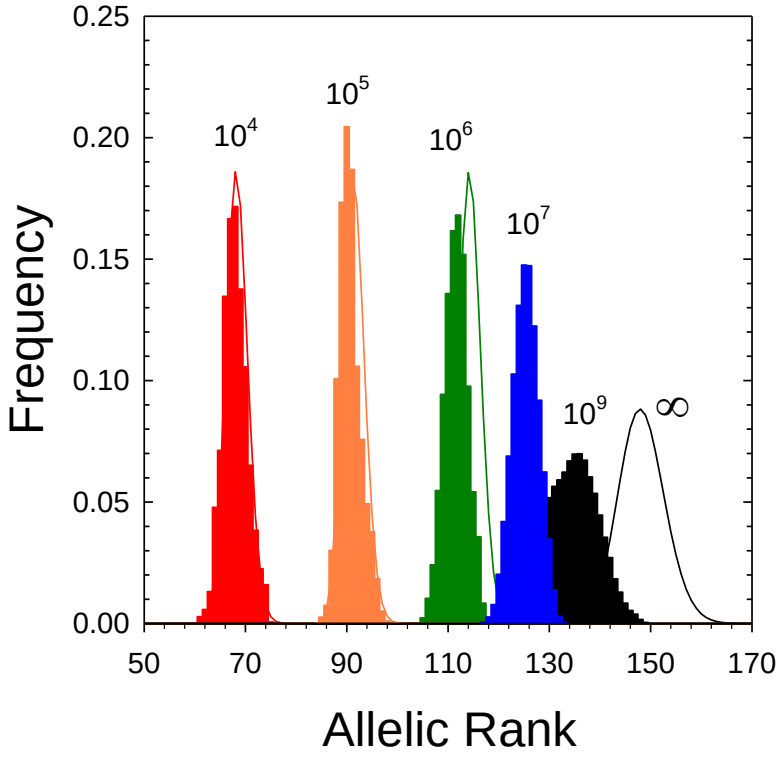

BIOGENESIS OF
TRANSCRIPTION MACHINERY
RNA polymerases
Spliceosomes
BIOGENESIS OF
TRANSLATION MACHINERY
Amino-acyl synthetases
Transfer RNAs
Ribosomes
TRANSCRIPTION
Base-loading fidelity
Splicing
TRANSLATION
Amino-acyl synthetase charging
Transfer RNA loading
Codon recognition
Messenger RNA surveillance
PROTEIN MATURATION
Folding
Post-translational modification
Assembly of subunits
A Nested Multitude of Cellular
Surveillance Mechanisms

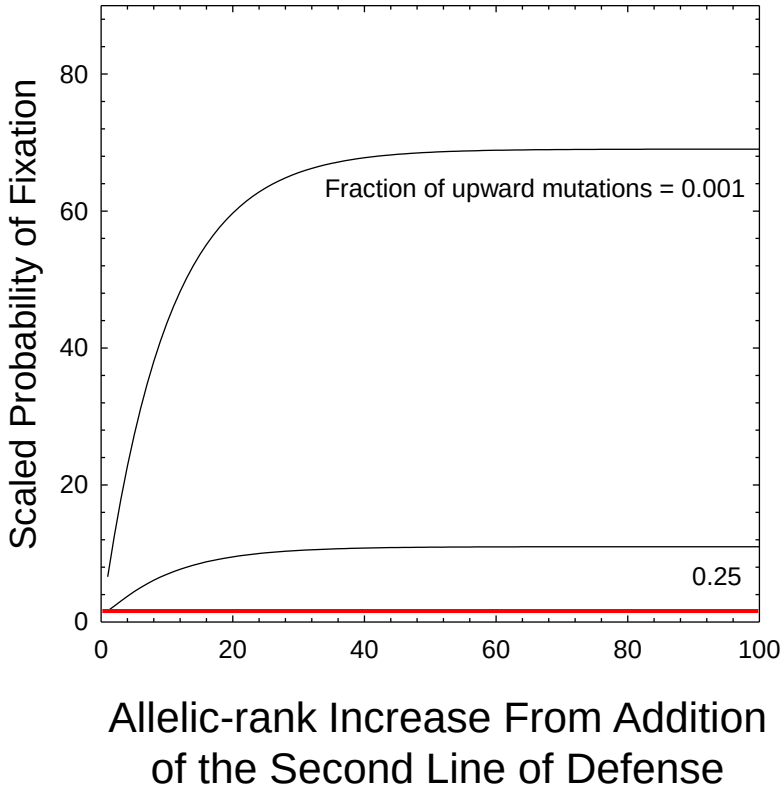
Evolutionary Layering and the Limits to Molecular Perfection:

1) Can a secondary layer of defense be added that breaks the drift barrier?

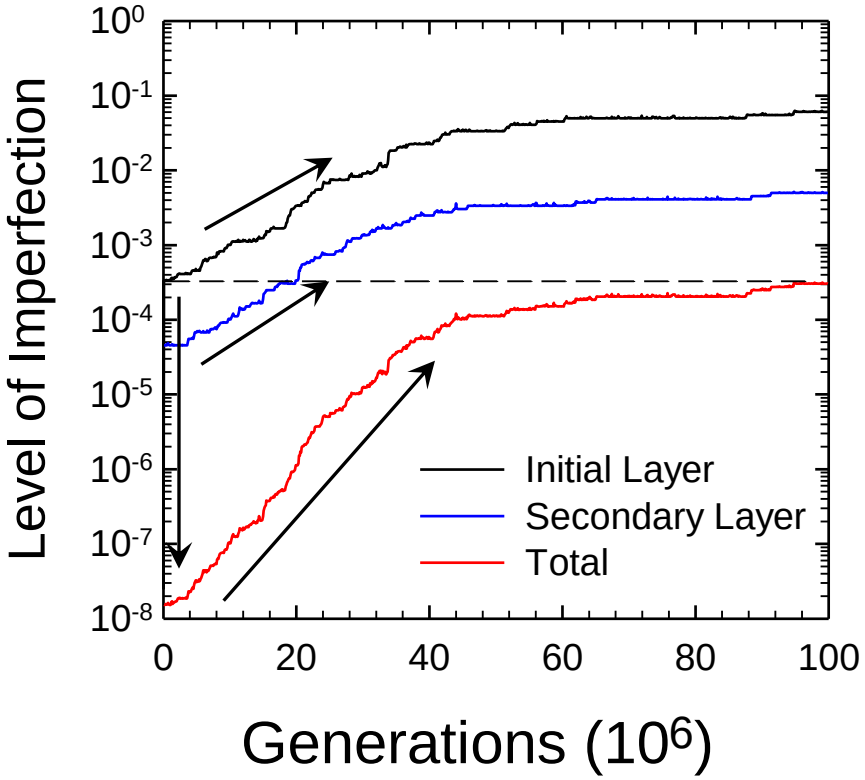
2) If such a genomic addition is assimilated, what are the long-term consequences for the previous layer, the new layer, and the combined effects of both?

Scaled Probability of Establishment of a Secondary Layer of Surveillance (relative to the neutral expectation)





The Fitness Boost From the Addition of a Layer of Accuracy Is Transient



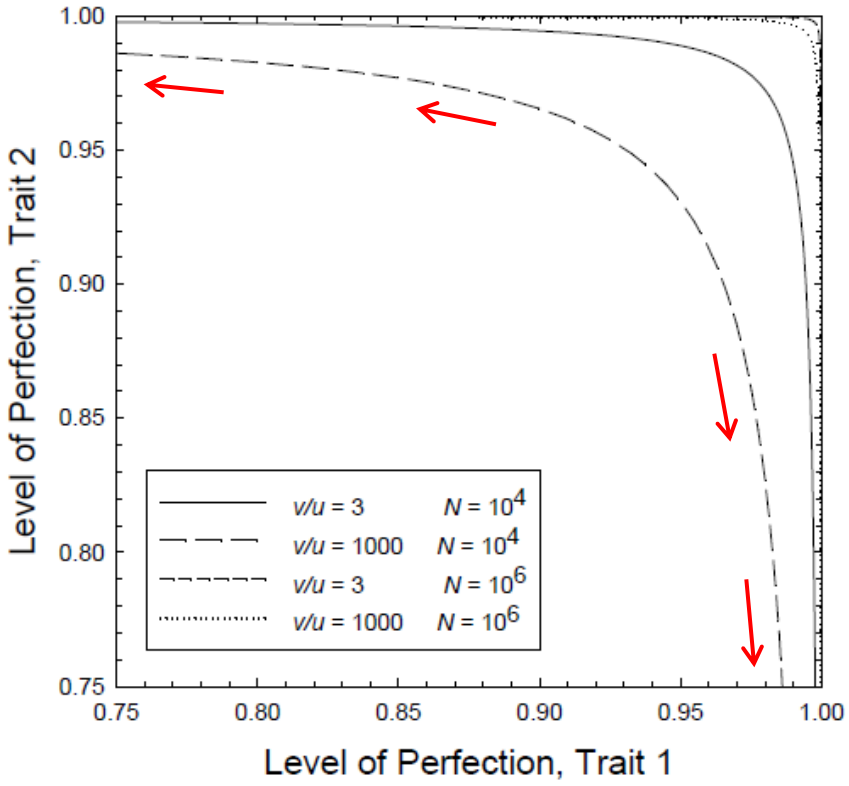
- Rapid improvement accompanies establishment of a new layer of protection.
- Both layers then gradually become less efficient.
- The level of overall performance returns to that for the single-layered state.

- The “Paradox of Robustness” (S. Frank, PLoS One): a more complex system evolves, but nothing is gained in the long run.
- Something has been lost: sensitivity of the system to mutational breakdown has increased.

A Bivariate Drift Barrier:

• Selection will operate to drive the joint effects of two traits down to the limits imposed by drift.

• There is a ridge along which the population can freely drift, even to the extent of losing one trait.



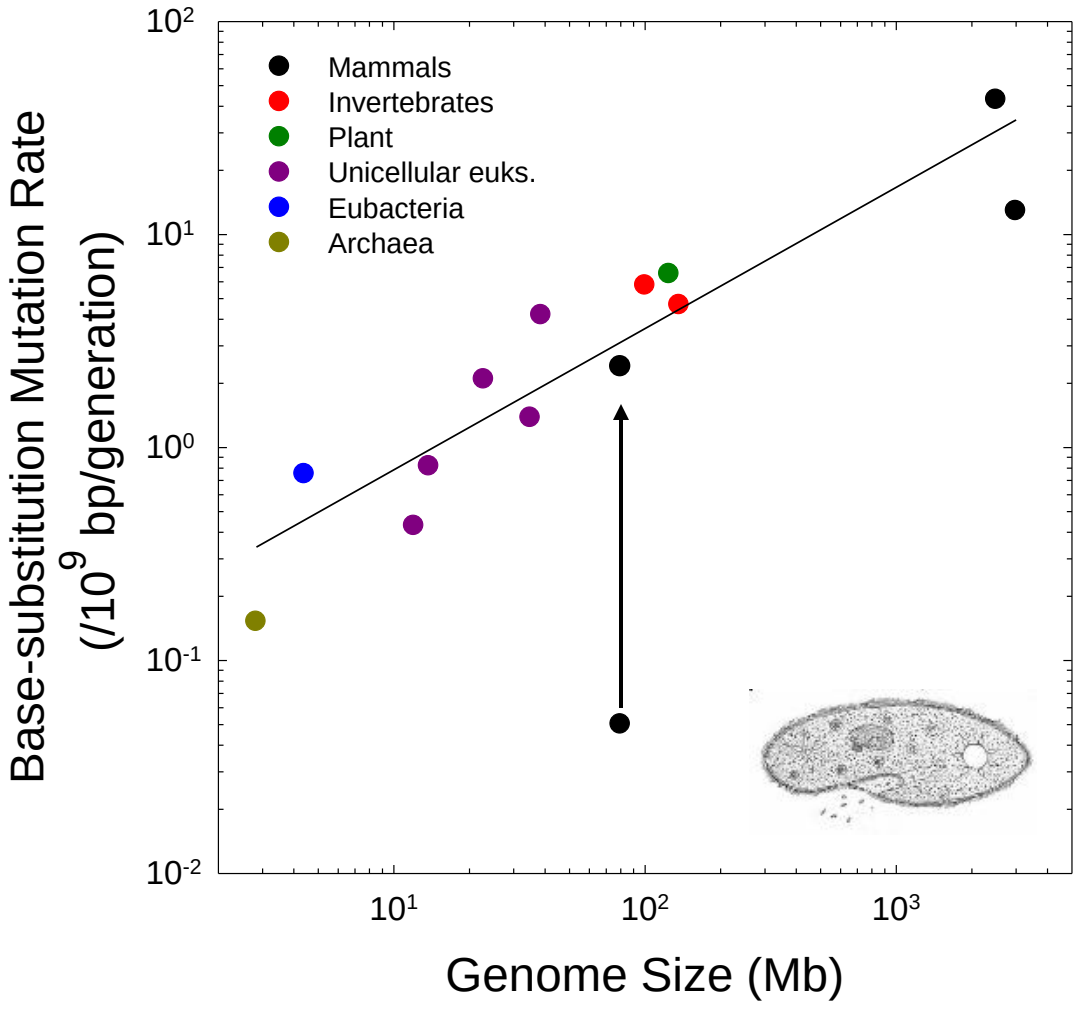
Some potential cell biological examples of transient redundancy resulting from a bivariate drift barrier:

- **Regulation of the cell cycle**: although the restriction of licensing of DNA replication origins to one event per cell cycle is critical to maintaining genome integrity, there is substantial variation within and among eukaryotic lineages in the mechanisms regulating such behavior (Drury and Diffley 2009; Cross et al. 2011).

- **Amino-acid biosynthetic pathways**: variations of pathway components exist within and among phylogenetic lineages, in some cases (e.g., lysine) with completely different pathways (Voet and Voet 2010).

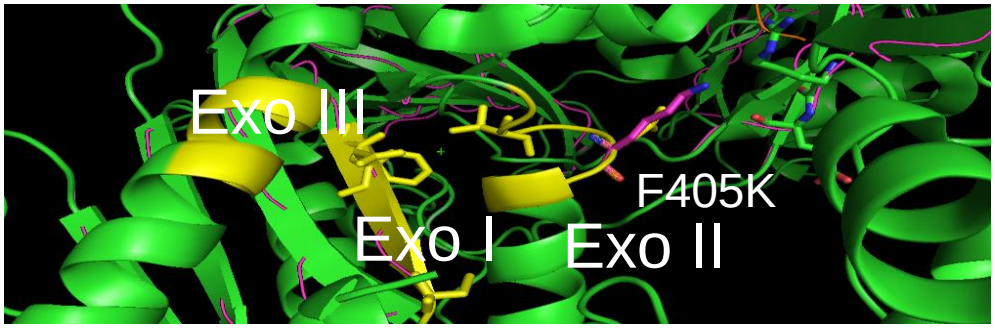
- **Protein folding**: in *E. coli*, genes whose protein products are clients of the molecular chaperone GroEL harbor significantly lower frequencies of optimal codons (and hence experience higher rates of misfolding associated with translational errors; Drummond and Wilke 2008) than do sporadic clients (Warnecke and Hurst 2010).

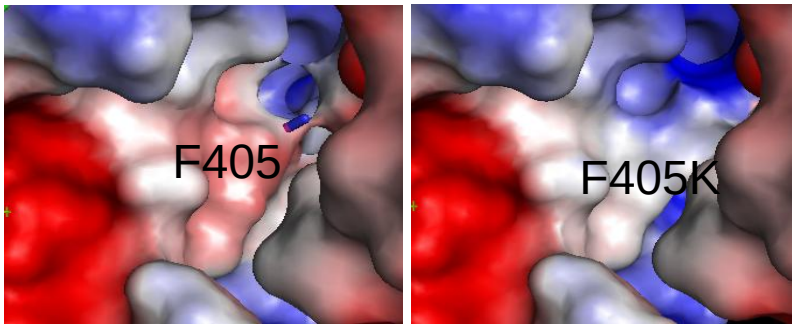
Paramecium Has the Lowest Known Mutation Rate Per Cell Division, Although Its Rate Per Sexual Episode is Compatible With Other Species



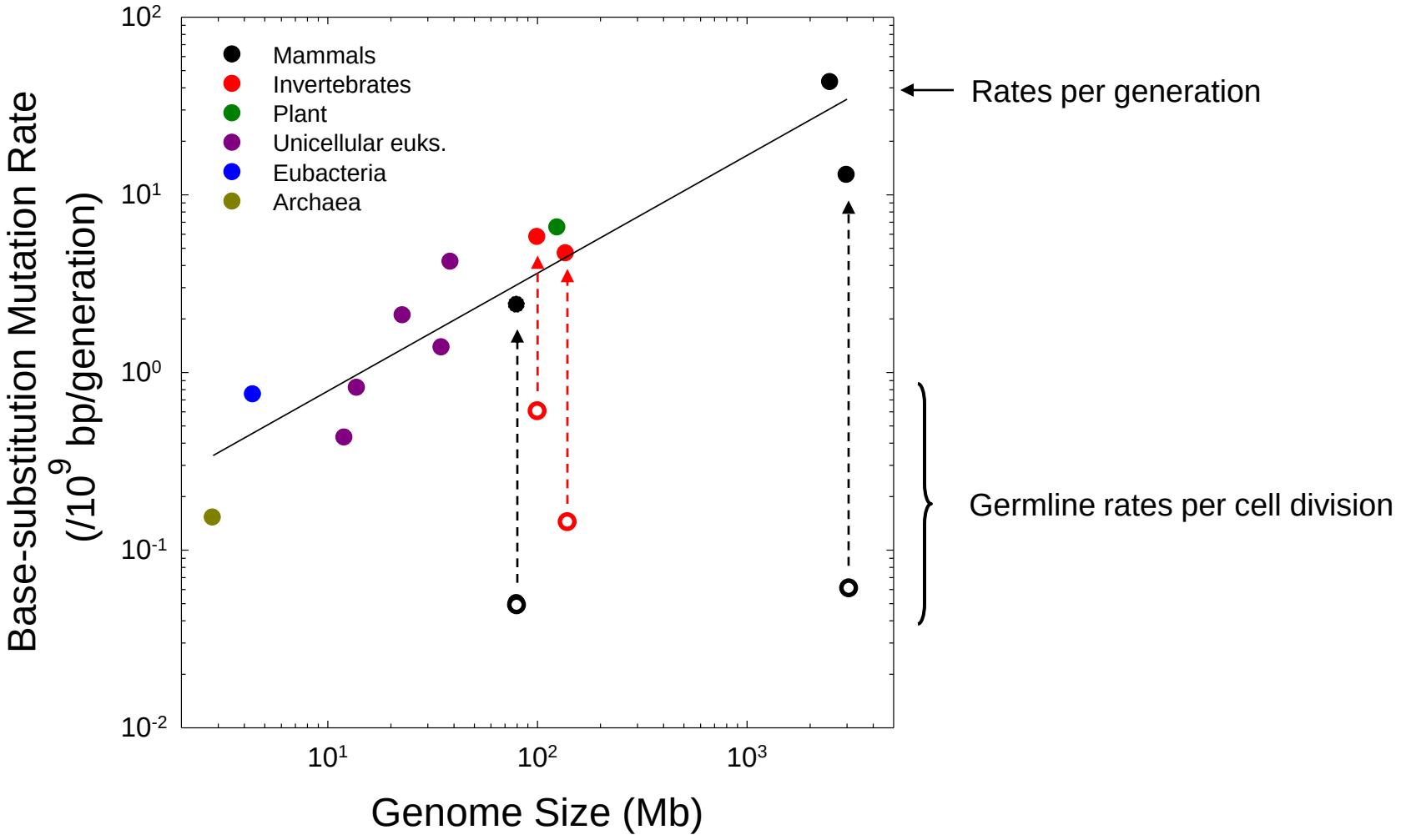
Some of the Major Replicative Polymerases in Ciliates Exhibit Radical Amino-acid Substitutions At Sites That Are Highly Conserved In Other Eukaryotes

Species	Polymerase Catalytic Sites: Region II	Region III	Region I	Proofreading 3' -> 5' exonucleases: Exo I	Exo II	Exo III
Epsilon - ε		*			+	*
H. sapiens	D V G A M Y P N I	K C I L N S F Y G Y	L E L D T D G I	F D I E T	N G D F F D	Y S V S D A
M. musculus	D V G A M Y P N I	K C I L N S F Y G Y	L E L D T D G I	F D I E T	N G D F F D	Y S V S D A
D. melanogaster	D V G A M Y P N I	K C I L N S F Y G Y	L E L D T D G I	F D I E T	N G D F F D	Y S V S D A
C. elegans	D V G A M Y P N I	K C I L N S F Y G Y	L E L D T D G I	F D I E T	N G D F F D	Y S V S D A
S. cerevisiae	D V A S M Y P N I	K V I L N S F Y G Y	L E L D T D G I	F D I E T	N G D F F D	Y S V S D A
A. thaliana	D V A A M Y P N I	K C I L N S F Y G Y	L E L D T D G I	F D I E T	N G D F F D	Y S V S D A
T. cruzi	D V G A M Y P N I	K C I L N S F Y G Y	L E L D T D G I	F D I E T	N G D Y F D	Y S V S D A
T. pseudonana	D V G A M Y P N I	K C I L N S F Y G Y	L E L D T D G I	F D I E C	N G D F F D	Y S V S D A
T. thermophila	D V A A M Y P N I	K I I L N S F Y G Y	L E L D T D G I	F D I E T	N G D K F D	Y S V S D A
P. tetraurelia	D V A A M Y P N I	K I I L N S F Y G Y	L E L D T D G I	F D I E T	N G D R F D	Y S I S D S
G. lamblia	D V A A M Y P N I	K C I L N S F Y G Y	L E L D T D G I	Y D I E T	N G D F F D	Y S V S D A
T. gondii	D V S A M Y P N I	K C I L N S F Y G Y	M E L D T D G I	W D I E C	N G D T F D	Y S V S D A

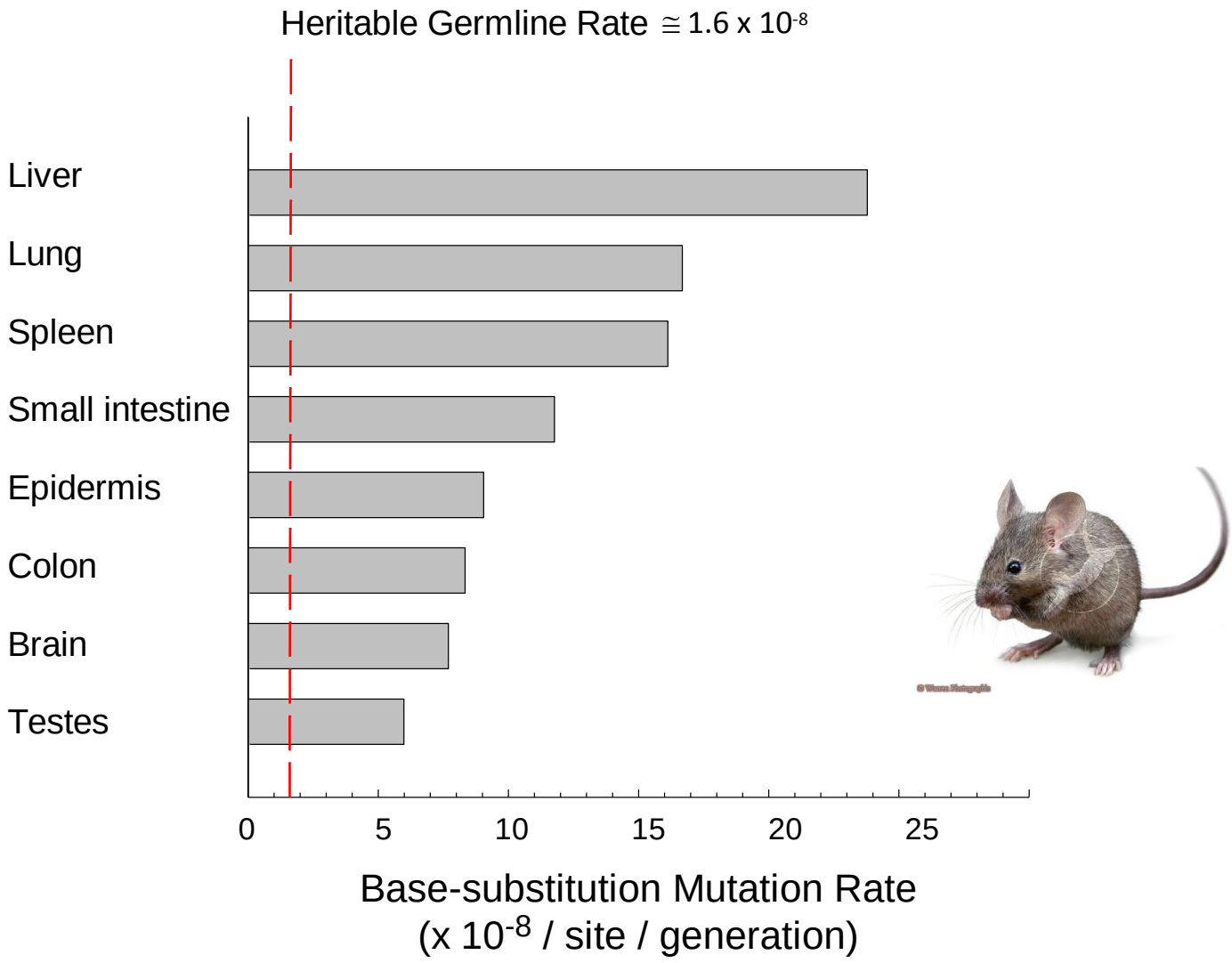




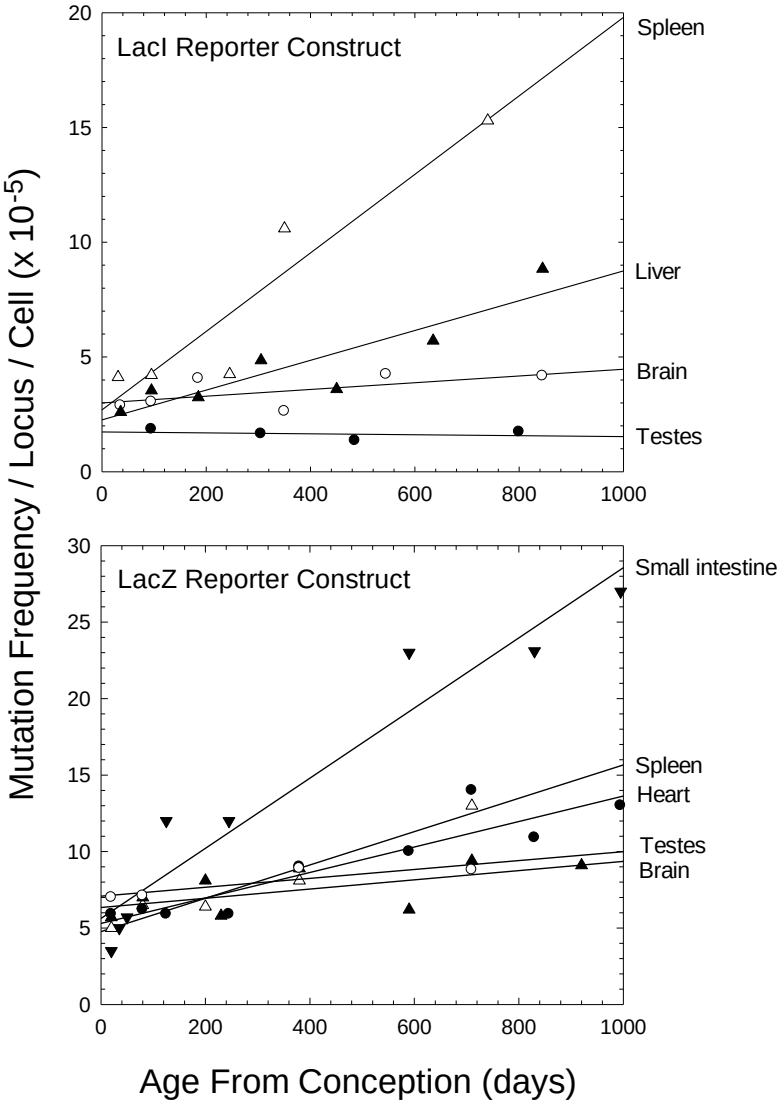
Mutation Rates Per Germline Cell Division in Multicellular Species Are Reduced To Levels That Accommodate The General Per-generation Pattern



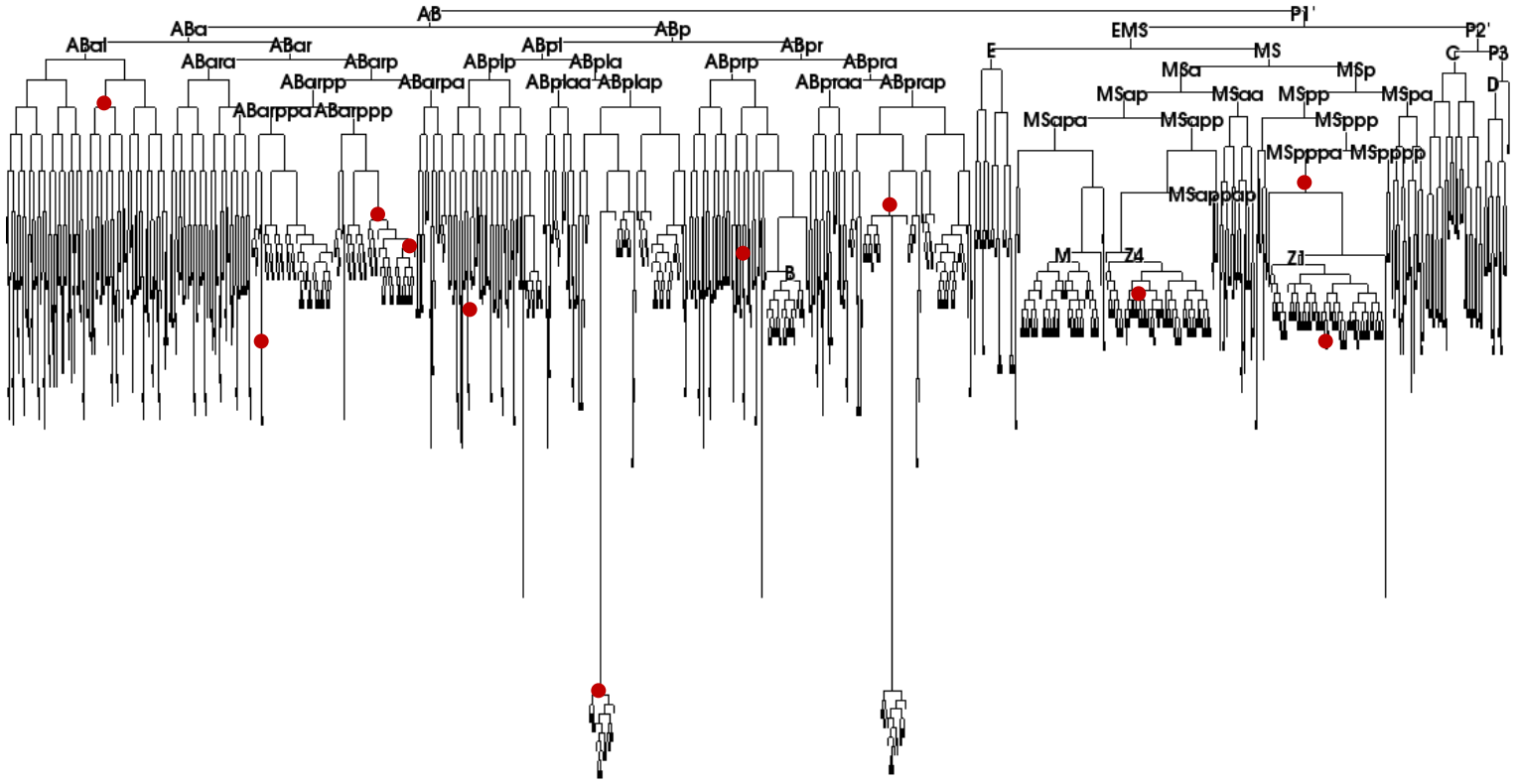
Mutation Rates in Somatic Tissues Are Up to 15x Those in the Germline



Somatic Mutations Accumulate With Age, But Only Weakly in the Germline



The Complete Cell Lineage of *Caenorhabditis elegans*



- Expected number of somatic mutations in an adult worm ≈ 10^3 cells x 10^8 bp/cell x 10^{-7} mutations/bp = 10^4

- Expected number for a human ≈ 10^{14} cell divisions x 10^9 bp/cell x 10^{-7} mutations/bp = 10^{16}/ soma
- Every cell will contain multiple mutations.

The Joint Indirect and Direct Fitness Effects of a Mutator Allele in a Multicellular Species

Selective disadvantage of an increment in the mutation rate =

Indirect heritable germline mutation load ($2 \times \Delta U \times s$)

+

Direct somatic effect

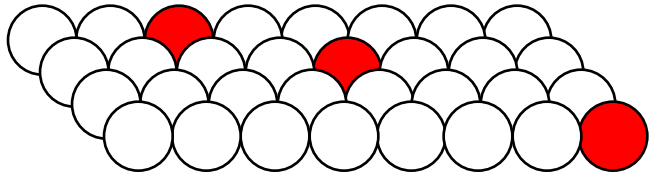
- Negative → cost of replication fidelity
- Positive → cost of somatic mutations

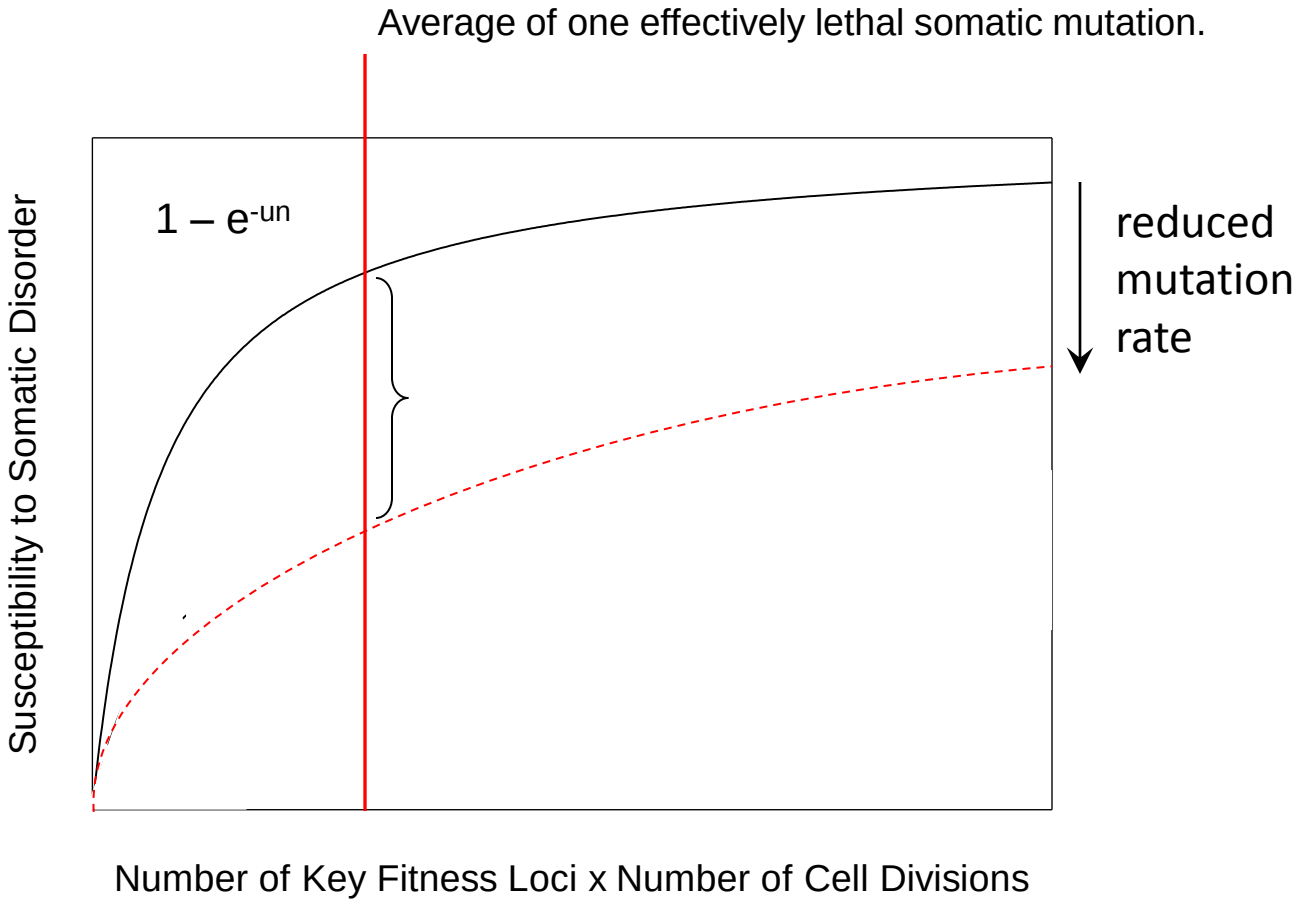
Somatic mutation will be an influential force in the evolution of the germline mutation rate if:

1) the same machinery is used in replication/repair in both kinds of tissue;

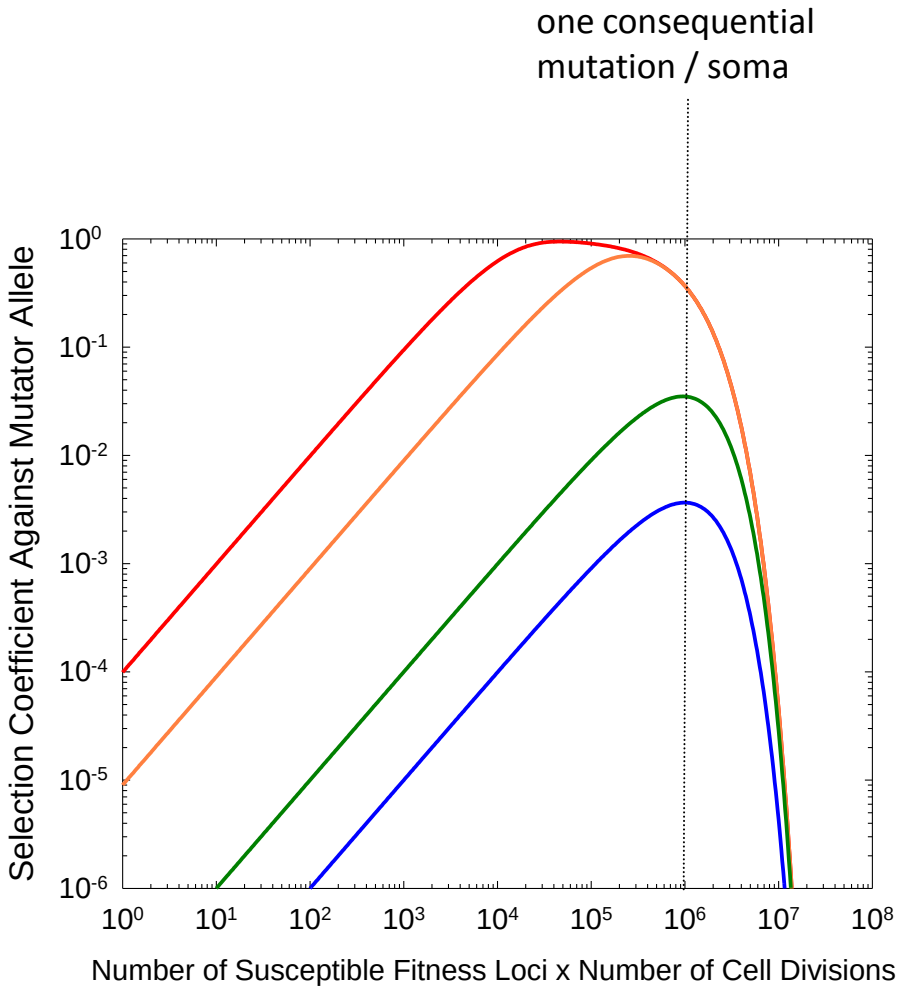
2) the somatic mutation load is on the order of the heritable germline mutation load ($2 \times \Delta U \times s$), or larger.

Multicellularity indirectly imposes selection pressure for a reduced mutation rate in a nonlinear manner.





Although the absolute magnitude of somatic mutation increases with the level of multicellularity, the relative selective disadvantage of a mutator allele decreases above a critical number of cell divisions.



MM = nonmutator

Mm = mutator heterozygote

Mutation rates:

$u_{MM} = 10^{-6}$ / allele / cell division

- $u_{Mm} = 100u_{MM}$
- $u_{Mm} = 10u_{MM}$
- $u_{Mm} = 1.1u_{MM}$
- $u_{Mm} = 1.01u_{MM}$

Major contributors:

Indiana University: Matthew Ackerman, Jamie Choi, Nicole Coffey, Tom Doak, Yana Eglit, Pat Foster, Matthew Hahn, Ignasi Lucas, Rohan Maddamsetti, Sam Miller, Sarah Schaack, Dan Schrider, Amanda Seyfert, Way Sung, Haixu Tang, Abe Tucker, Emily Williams.

University of New Hampshire: Kelley Thomas, Vaughn Cooper, Shilpa Kulkarni, Krystalynne Morris, Kazufusa Okamoto.

University of Utah: Joe Dickinson.

Max Planck Institute, Tubingen: Detlef Weigel, Korbinian Schneeberger, Stephan Ossowski, Norman Warthmann.

Harvard University: Daniel Hartl, Christian Landry, Eric Dopman.

University of Florida: Charles Baer.

Florida State University: David Houle

Oregon State University: Dee Denver.

University of Minnesota: Ruth Shaw.

The Unique Genomic Landscape in Humans

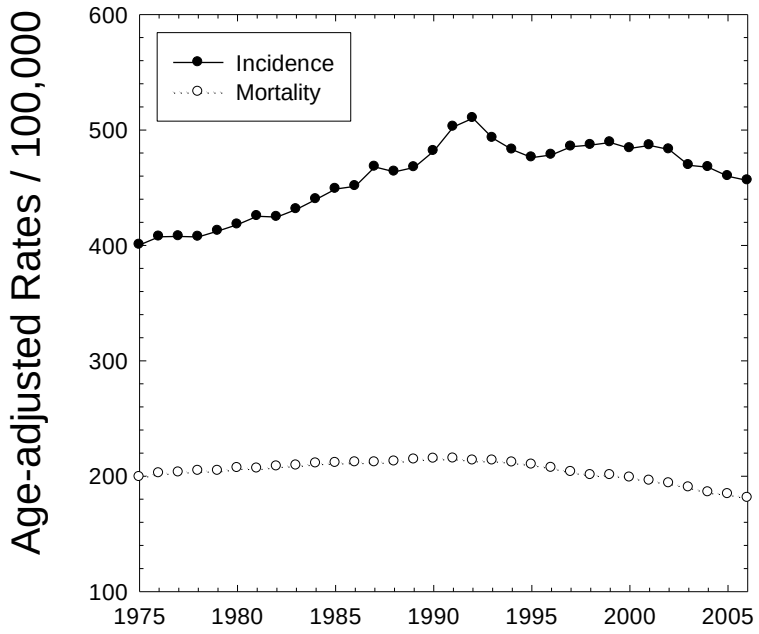
- The extreme population-genetic environment of humans (high power of drift, low power of recombination, and high power of mutation) magnifies the ability of mutator alleles and mutationally hazardous DNA to accumulate in an *effectively neutral* fashion.

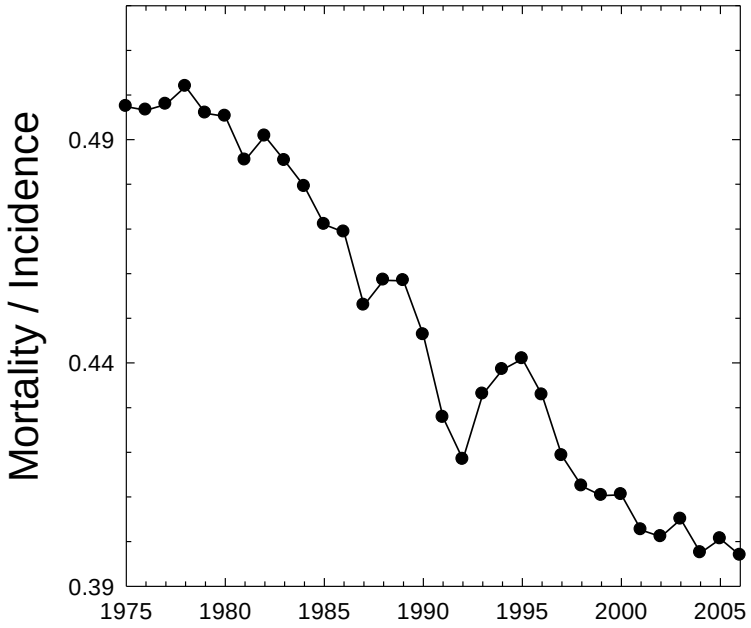
- Once multicellularity has reached an extreme form, the power of selection against somatic genetic disorders (cancer) is decreased.

- Humans are uniquely capable of behaviorally determining the fates of deleterious alleles – modern medicine effectively encourages the buildup of mutational load.

H. J. Muller

All Cancers in the US Population



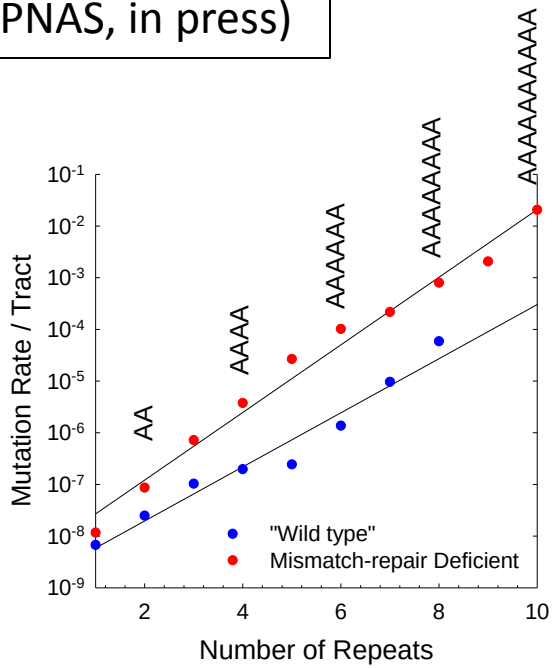


National Cancer Institute, Statistics Review, 2009

Spatial Variation in the Mutation Rate (P. Foster et al., PNAS, in press)

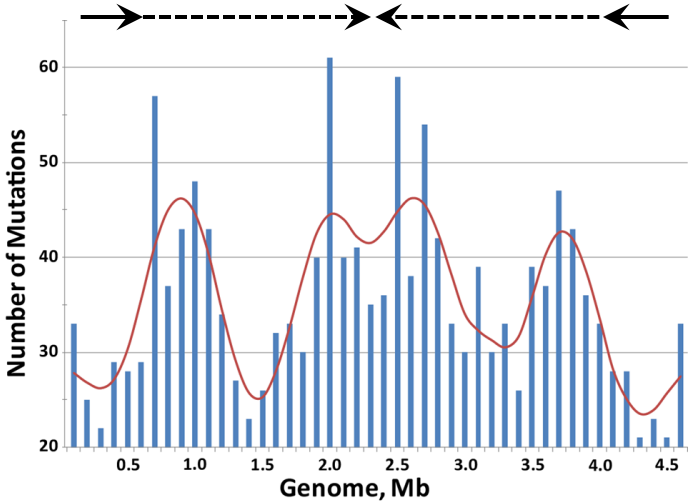
Strain fingerprinting with hypervariable positions.

- The mutation rate to length variants dramatically increases with the length of homopolymeric runs.



The mutability of a gene depends on its chromosomal location.

- Mutations are distributed across the genome in a large-scale, periodic pattern, repeated in mirror-image in each half of the genome.



The Paradox of Universal Health Care / Personalized Medicine

- The human imperative is to magnify the probability of survival and reproduction regardless of the level of genetic affliction.

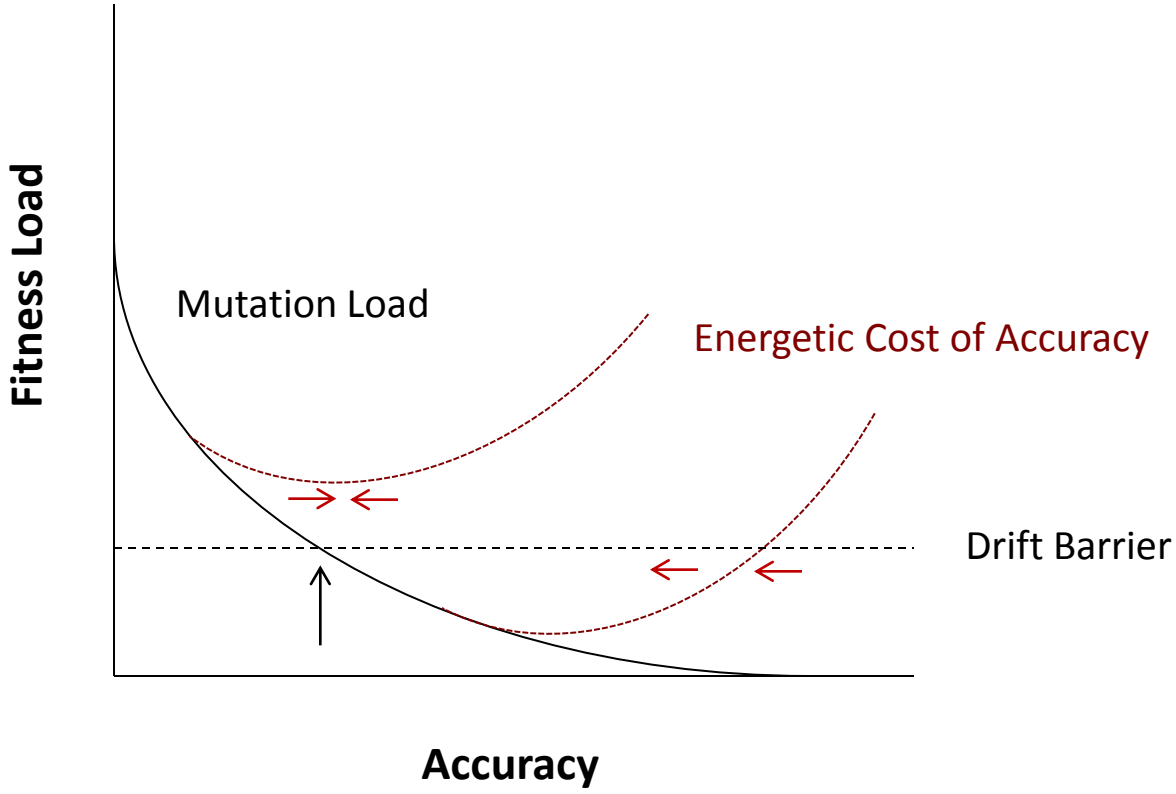
- At least one to two deleterious mutations arise per human genome per generation.

- The average deleterious effect of such mutations is very mild, ~1 to 2.5% per event.

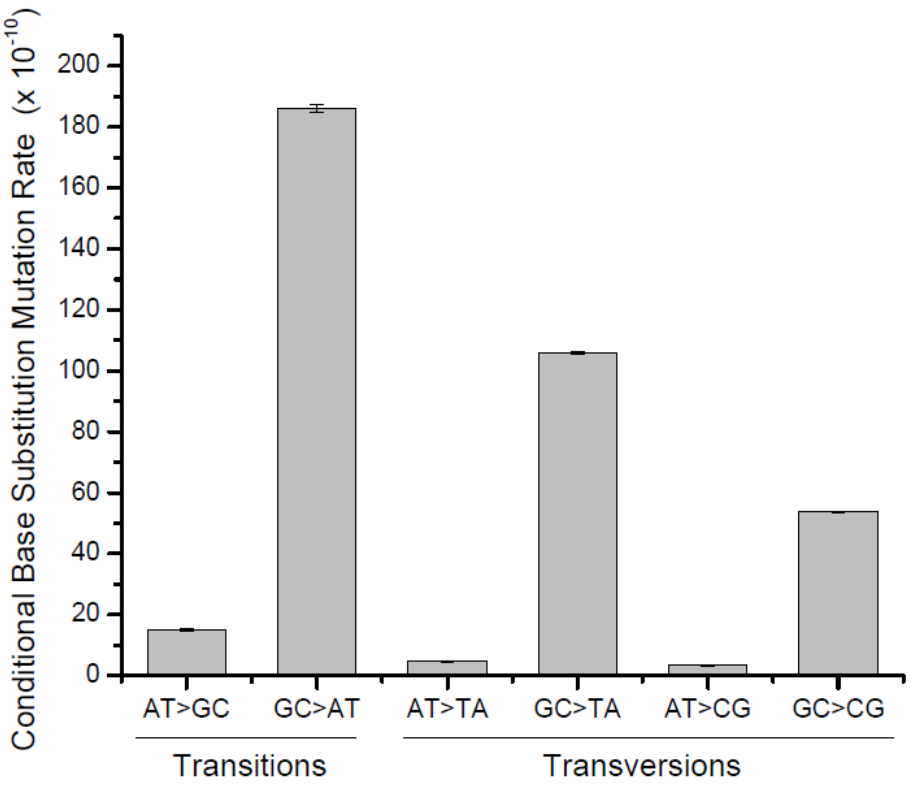
- With a complete relaxation of selection, the decline in fitness per generation is 1 to 5% per generation, or 3 to 15% per century.

- This rate of decline in human fitness operates on a time scale comparable to global warming.

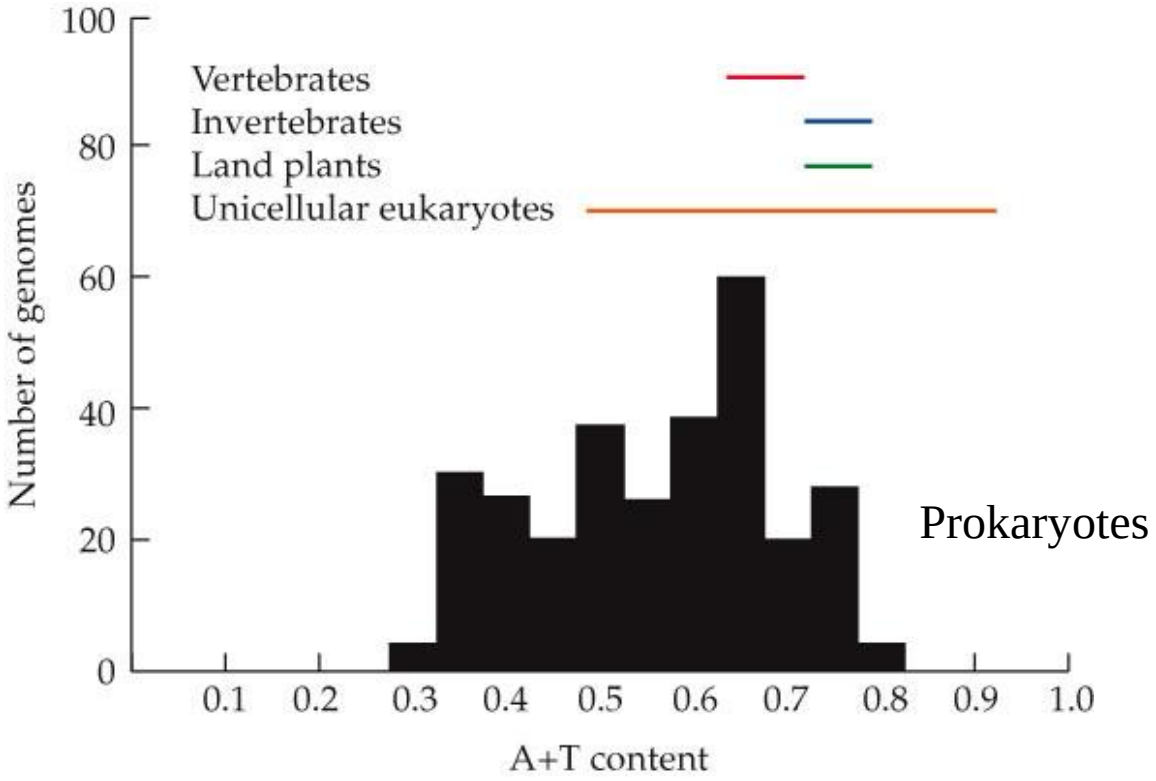
The Limits to Molecular Perfection: the Drift-barrier Hypothesis or Physical Constraints?



A Strong Mutational Bias Towards A/T Production in *Mesoplasma florum*



What is the source of the wide phylogenetic range of variation in nucleotide usage?



- All genomes have substantial mutation bias towards A/T production.
- Genome-wide nucleotide compositions are not in mutation equilibrium.
- The universal genomic deficit of A/T must be a result of selection and/or biased gene conversion.